YALE ECONOMIC ESSAYS

VOLUME 2 NUMBER 1

An Empirical Examination of a Commercial Bank Loan Offer Function — *Donald D. Hester*

Investment Criteria and Choice of Techniques of Production — *Thirukodikaval N. Srinivasan*

An Empirical Study of Cash, Securities, and Other Current Accounts of Large Corporations — *Alan W. Heston*

Speculation in the Foreign Exchanges: The European Experience, 1919-1926 — *Robert Z. Aliber*

Long-Run Adjustments to Changes in the Capital-Output Ratio and the National Debt — *Richard S. Thorn*

SPRING 1962

YALE ECONOMIC ESSAYS

This journal is intended to serve two principal functions. All doctoral dissertations in economics at Yale University will be published in its pages either in full or in extended summary. It is also a medium for economic comment and analysis by faculty, students, and graduates in economics of Yale University.

Editorial Board: Neil W. Chamberlain
Raymond P. Powell Arthur M. Okun

Editorial Staff: Bruce W. Morgan and Karen H. Hester

Volume 2 Spring 1962 Number 1

Yale Economic Essays is printed at the Carl Purington Rollins Printing-Office of the Yale University Press. A one-year subscription (two issues) is $4.00. Single copies are $2.50. The rate for students for one year is $2.00. Checks should be made payable to Yale University. Please address any correspondence to the Editorial Board, 37 Hillhouse Avenue, New Haven, Connecticut.

Copyright 1962 by the Yale Economic Essays.

YALE ECONOMIC ESSAYS

TABLE OF CONTENTS

An Empirical Examination of a Commercial Bank Loan Offer Function — *Donald D. Hester* — 3

Investment Criteria and Choice of Techniques of Production — *Thirukodikaval N. Srinivasan* — 59

An Empirical Study of Cash, Securities, and Other Current Accounts of Large Corporations — *Alan W. Heston* — 117

Speculation in the Foreign Exchanges: The European Experience, 1919–1926 — *Robert Z. Aliber* — 171

Long-Run Adjustments to Changes in the Capital-Output Ratio and the National Debt — *Richard S. Thorn* — 247

Summary: This paper proposes a theory to explain the terms at which a commercial bank will lend to a business loan applicant with particular financial characteristics. Some hypotheses from the theory are subjected to empirical test. Four terms of lending are analyzed in considerable detail: the rate of interest on the loan, its maturity, its amount, and the existence of security.

Empirical tests are performed on three samples of term loans collected from large commercial banks and on two samples from the 1955 and 1957 business loan surveys conducted by the Board of Governors of the Federal Reserve System. Among the characteristics of the loan applicant which are found to influence terms of lending are his profits, profit rate, current ratio, total assets, demand deposit balances, location, and the length of the period during which he had held deposits with the lending bank.

Deposit instability of the lending bank and the proportion of loans in its portfolio are also hypothesized to influence a bank's willingness to lend. Empirical tests verify these hypotheses.

The "availability of credit" doctrine is interpreted in terms of the theory and subjected to empirical test. A conclusion of the test is that there is little empirical evidence of an increase in credit rationing during the 1955–1957 period, contrary to this doctrine.

Statistical tools employed in the analysis include analysis of variance, multiple regression, and canonical correlation.

This essay was submitted to Yale University in 1961 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is currently Assistant Professor of Economics at Yale University and on the staff of the Cowles Foundation for Research in Economics at Yale.

Dissertation Committee: Professors Arthur M. Okun (Chairman), James Tobin, Raymond P. Powell, and Henry C. Wallich.

*

I am grateful to Professors Arthur Okun, Raymond Powell, Henry Wallich and James Tobin, and to T. N. Srinivasan and Karen Hester for many helpful comments. This research was generously supported by the American Bankers Association and the Cowles Foundation for Research in Economics at Yale. I assume responsibility for all errors remaining in this paper.

D. D. H.

AN EMPIRICAL EXAMINATION OF A COMMERCIAL BANK LOAN OFFER FUNCTION

Donald D. Hester

1. Theory of the Loan Offer Function

On June 29, 1960, commercial banks in the United States had approximately 115 billion dollars of loans outstanding and were the largest source of short term funds for business. These credits were extended to a wide variety of borrowers and the terms of loans varied considerably among borrowers. Despite the obvious importance of bank lending, economists have not concerned themselves with a theory to explain this activity in detail. The present paper proposes a theory of bank lending and introduces supporting statistical evidence. Implications of the theory for monetary policy are considered in the final section.

This paper considers only the terms at which loan applicants may obtain loans from commercial banks. It abstracts from bank choices among bonds, loans, and cash, except to the extent that these choices appear to influence the terms at which a bank will lend.[1] Services of banks as consultants or sources of information are ignored.[2] The deposit structure, equity, management, etc., of any particular bank are assumed to be fixed. The loan market is assumed to be imperfect.

A loan offer function is a relation which specifies the terms at which a bank with particular characteristics is willing to lend to a borrower with a known profit, balance sheet, and credit history and with particular prospects for the future. It is a generalized supply function for loans in the sense that, instead of merely having the amount of loans determined by a set of exogenous variables, it has a set of loan terms including the amount of loans determined by the set of exogenous variables.

Any loan application may be characterized by a set of terms of lending, $\underline{t} \in \underline{T}$, where $\underline{T}$ is the set of all imaginable sets of terms of lending. Thus $\underline{t}$ might refer to a loan of one million dollars, possessing a maturity of one year, payable quarterly, having an interest rate of five percent, etc. A potential borrower has balance sheets, income statements, and a credit rating which will be denoted by the set $\underline{v} \in \underline{V}$, where $\underline{V}$ is the set of all imaginable sets of information about borrowers. A lending bank has a set of characteristics, $\underline{u} \in \underline{U}$, where $\underline{U}$ is the set of all imaginable sets of characteristics of lending banks. Thus $\underline{u}$ might include a bank's deposit instability, its loan

---

1. Cf. Richard C. Porter, "A Model of Bank Portfolio Selection," Yale Economic Essays, I (Fall 1961), pp. 323-359; Roland I. Robinson, The Management of Bank Funds (New York: McGraw-Hill Book Company, Inc., 1951); and David Chambers and Abraham Charnes, Inter-temporal Analysis and Optimization of Bank Portfolios (O. N. R. Research Memorandum No. 27, The Technological Institute, Northwestern University, 1960).

2. Cf. George Katona, Business Looks at Banks (Ann Arbor: University of Michigan Press, 1957).

experience, the present state of its portfolio, rates of interest on competing assets available to it, etc. Abstractly, a loan offer function may be expressed:[1]

1 - 1) $F(t; u, v) = 0$.

Section 1 develops a theory of bank lending in terms of expression 1 - 1. Sections 2, 3, and 4 present empirical evidence on this theory. Section 5 summarizes the analysis and suggests its policy implications.

In Parts 1 - 1 through 1 - 3, the discussion is considered to be suggestive rather than exhaustive. Other terms of lending, characteristics of borrowers, and characteristics of banks can be imagined. Similarly, quotations are intended to be illustrative rather than an exhaustive study of opinion in the banking literature.[2] Finally, the argument in these sections frequently is derived from interviews which I held with numerous bank officers.

1 - 1. Analysis of Terms of Lending

It is assumed throughout 1 - 1 that a loan applicant summarized by $\underline{\bar{v}} \in \underline{V}$ applies for a loan from a bank summarized by $\underline{\bar{u}} \in \underline{U}$. It is convenient to introduce the analysis of terms of lending by considering a relation implicit in the production function, the production possibility curve.

A production possibility curve may be defined to be the efficient set of outputs obtainable from a given set of inputs. The co-ordinates of this curve are measured in units of the goods which could possibly be produced. In the case of the loan offer function, the analogue of this curve is the efficient set of terms $t^* \in T'$ obtainable from given sets of characteristics, $\underline{\bar{u}}$ and $\underline{\bar{v}}$.[3] The co-ordinates are measured in units which are applicable to each term; thus maturity might be measured in months.

Efficiency in the sense of Pareto optimality implies that no more output of any good can be obtained from a particular set of inputs without reducing the output of some other good. Similarly, efficiency in the case of the analogous loan offer curve implies that a bank will not grant a more liberal loan from the point of view of one term or condition without

---

1. Lawrence Ritter, in an interview, brought my attention to a manuscript by J. M. Guttentag. Mr. Guttentag's paper contained a description of bank lending behavior which was somewhat similar to the loan offer function. Although I had already developed the notion of a loan offer function when I first saw his manuscript, it is possible that my subsequent thinking was influenced by his draft. He presented no detailed statistical evidence. Cf. J. M. Guttentag, "Credit Availability, Interest Rates, and Monetary Policy," The Southern Economic Journal, XXVI (January 1960), pp. 219–228.

2. Despite the prescriptive appearance of some of the quotations, it is believed that the authors intended their statements to be essentially descriptive.

3. $T' \in T$. T' is the obtainable set of terms of lending.

worsening other terms. The notion of liberal terms of lending is considered below.

Terms of loans include the size; rate of interest; maturity; schedule of repayments; guarantees; other collateral; restrictions on dividends, working capital, salary increases, plant and equipment acquisitions, retained earnings, and personnel loans; and subordination of outstanding debts. Other terms include commitment fees charged by the bank, costs of audits, penalties for early repayments which come from other borrowing, verbal agreements on future balances, etc. Space considerations limit the discussion to four principal terms which are analyzed empirically in Sections 2-4. These are the rate of interest, the size of the loan, the maturity of the loan, and whether or not the loan is secured.

The following statements are conjectures about a bank's view of each of the four terms of lending, assuming the remaining three terms are established. If a term is more liberal, it is less preferable from the point of view of the bank.

Obviously loans which are made at higher rates of interest are preferred by banks. It is assumed that a bank's marginal utility of income is always positive. [1]

Banks also prefer relatively short maturity loans. This is partly attributable to the potential short term character of their demand deposit liabilities. Further, banks view default risk as a function of exposure time (maturity). [2]

Banks do not desire to make extremely short maturity loans (say one hour). The costs of investigating borrowers cannot ordinarily be covered by the interest on the loan. However, this is not an important qualification because few borrowers desire one hour loans. [3]

---

1. Donald Hodgman has observed that higher loan rates of interest may not be preferable if they increase the probability of default or only partial repayment by a borrower. This result hinges on an assumption that interest payments on a loan are not collected continuously. No contradiction exists between Hodgman's argument and the statement in the text. It is quite possible that banks prefer not to lend to a poor risk when interest rates rise. It is only argued that if a loan is to be granted, banks will be made better off (or no worse off) by increasing the loan rate of interest. Cf. Donald Hodgman, "Credit Risk and Credit Rationing," The Quarterly Journal of Economics, LXXIV (May 1960), pp. 258-278.

2. "For any given credit rating, a longer term loan should have a higher rate if for no other reason than that the liquidity differential in favor of short-term credits is bound to be worth something. The greater errors which are bound to creep into long-term credit judgments merit some differential." Robinson, op. cit., p. 197.

3. This paragraph ignores special arrangements like loans to brokers and security dealers.

Banks prefer to lend moderate amounts rather than very large or small sums. Small loans have high handling charges relative to the interest on the loan. The analysis is restricted to loans where handling costs are relatively small.[1]

Large loans are more risky in two ways. First, a borrower who obtains a large loan will have a higher debt-asset ratio after the loan has been consummated and thus will have a larger default risk. Second, banks will not be able to diversify their loan portfolio optimally if equal-risk borrowers get loans of different amounts.[2]

If all other terms of lending are specified and if borrowers are charged the costs of providing collateral, banks prefer to make secured loans. Collateral reduces the default risk of the loan. Frequently borrowers do bear the costs of providing collateral.

In conclusion, bankers attempt to make loans which have high interest rates, short maturities, moderate amounts, and which are secured. Borrowers attempt to get more favorable terms by seeking low rates of interest, no security, and frequently larger amounts and longer maturities. If borrowers are effective bargainers, banks will end up on their offer frontiers.[3] An important aspect of this description of lending is that both bankers and borrowers may alter any of the terms of lending during negotiations.[4]

Despite the restrictive assumptions which have been imposed to develop these notions, evidence of the relationships among terms of lending can be found in the banking literature:

> We find it difficult to do more than generalize on rate. It is of course elementary to say that the rate should vary with the risk and is affected by general and sectional money market conditions.

---

1. Roland I. Robinson places considerably more emphasis on the costs associated with bank lending. "A big borrower seeking a small accommodation can get a better rate than a small borrower, but he will have to pay more than he would have to pay for a larger sum, presumably still within the credit capacity of the borrower." Ibid., p. 119.

2. This statement assumes that the probability of an individual borrower defaulting on his loan is independent of the probabilities of default of other equal-risk borrowers. Cf. Harry M. Markowitz, Portfolio Selection (New York: John Wiley and Sons, Inc., 1959).

3. It is assumed that a bank's loan offer function is invariant with respect to the course of negotiations.

4. Henry Wallich observes that banks may adjust the terms of lending for certain classes of borrowers: "Moreover, the premium which the less-than-gilt-edged borrower must pay may take forms other than interest, such as an inconveniently short maturity, or surrender of a share of control." "The Changing Significance of the Interest Rate," The American Economic Review, LVI (December 1946), p. 766.

Correspondingly, it is elementary to say that the longer the period the higher the average rate should be.[1]

The use of security in extending business credit is a common practice among banks in the United States. The main purpose of collateral or other security is to limit the bank's risk of loss and facilitate collection in the event the borrower is unable or unwilling to repay the loan at maturity. While banks generally insist on collateral where circumstances do not justify their making an unsecured loan, they still depend for repayment in large part on the good faith of the borrower. In some cases where the bank would be willing to extend unsecured credit, the borrower may find it advantageous to use security to obtain a larger loan, a longer maturity, or a lower rate of interest.[2]

1 - 2. Analysis of Borrower Characteristics

It is assumed throughout 1 - 2 that a loan applicant requests a loan with terms $\underline{t} \in \underline{T}$ from a bank summarized by $\underline{u} \in \underline{U}$. It is again convenient to introduce the topic by first considering a derivative of the production function.

Isoquants of production functions show what minimum combinations of inputs to a production process are necessary to obtain a particular set of outputs. Borrower isoquants of a loan offer function show what minimum combinations of characteristics (if any) of a borrower, $\underline{v} \in \underline{V}$, are necessary in order that the bank be willing to lend at terms $\underline{t}$.

It is elementary to observe that banks desire information about prospective borrowers in order to reduce the likelihood of making loans which will not be repaid.[3] But what information should banks collect?

Clearly the kind of information should not be independent of the terms of the loan request. Very small loans or very short maturity loans do not generate sufficient revenue to offset expensive data collection and analysis. Long term loans require emphasis on future earnings rather

---

1. "Report of the Committee on Longer Term Loans," Bulletin of the Robert Morris Associates, Supplement, November 1937, p. 5.

2. Caroline H. Cagle, "Security Pledged on Business Loans at Member Banks," Federal Reserve Bulletin, XLV (September 1959), pp. 1115.

3. Roland I. Robinson has emphasized the importance of avoiding default risks: "The arithmetic of lending and loan interest is impressive. The average rate of return on member bank loans in 1948 was 3.8 percent. About two-thirds of this amount was absorbed in expenses leaving only slightly more than 1 percent for losses on loans and profits for stockholders. An average loss ratio of 1/2 of 1 percent would have meant that the profit rate was halved; it disappeared if average losses went to 1 percent of outstanding loans." Robinson, op. cit., p. 102.

than the current position of the firm.[1] Security may well obviate the necessity for a detailed credit investigation.

Nevertheless, it is possible to list information to which bankers frequently refer when judging loan applications. If applications for loans with different terms are received, the signs of the relationships between the various terms of lending and the following variables remain unchanged; only the weights placed on the variables may be altered. Important information about loan applicants includes the present and past size of his current assets, liquid assets, working capital, current ratio, inventories, total assets, outstanding debt, net worth, profits, sales, and deposit balances; the age of his inventories, plant, and equipment; the stability of demand for his products; and certain qualitative information, the purpose of the loan, the business of the borrower, the integrity and competence of management, the labor relations of the borrower, etc.[2]

Some indication of the significance of each piece of information can be inferred from standard works on statement analysis by Wall, Foulke, and Guthmann.[3] Although comprehensive systems of statement analysis have been developed, none is infallible. Different credit men do not place the same emphasis on each of the above or other pieces of information. The remainder of this part considers certain of the above variables which can easily be subjected to quantitative examination.

The expected after-tax profits of a loan applicant is a critical variable to banks for three reasons. First, a future stream of profits yields a cash flow which may be used to repay loans or meet future unforeseen emergencies. Second, the firm may grow as a result of the profitability of the business which implies that the bank may be able to obtain larger deposit balances and profitable loans in the future. Finally, banks may believe that profits are evidence of the competence of an applicant's management.

The current ratio, working capital, current assets, and liquid assets are all measures of a firm's ability to withstand immediate temporary

---

1. ". . . obviously, in general, short term credit is repaid from seasonal movements, while term loans depend upon realization of projected earnings in the majority of cases . . . ." "Term Loans," Bulletin of the Robert Morris Associates, Supplement, July 1957, p. 11.

2. Robinson, op. cit., Ch. 8, and Benjamin H. Beckhart, ed., Business Loans of American Commercial Banks (New York: The Ronald Press Company, 1959), Ch. 4.

3. Alexander Wall, How to Evaluate Financial Statements, 1st ed. (New York: Harper and Brothers Publishers, 1936); Roy A. Foulke, Practical Financial Statement Analysis, 1st ed. (New York: McGraw-Hill Book Company, Inc., 1945); and Harry G. Guthmann, The Analysis of Financial Statements, rev. ed. (New York: Prentice Hall, Inc., 1936).

adverse business fluctuations.[1] Different analysts place different emphasis on each of the measures. These variables are particularly important for loans of short maturity. Trends in these variables may reveal developing credit weaknesses.

Uncontrolled inventories are evidence of impending bankruptcy.[2] If an applicant's inventories are rapidly growing or particularly high, banks are cautious when extending credit. The ratio of inventories to sales for a firm relative to its industry is a measure of whether inventories are controlled. Over time the firm's ratio may reveal a highly cyclical character of its demand. Fluctuations in demand may prevent a borrower from repaying his loan on schedule.

Net worth is a measure of the solvency of a firm. Obviously default risks rise as loan applicants become less solvent. High ratios of debt to net worth may portend a decline in the firm's solvency at some future date.[3]

Finally, a bank is less willing to reject loan applications from firms carrying large deposit balances with it. From interviews with bankers, I inferred that they believe good present balances imply good future balances. In effect, firms carrying large idle balances with commercial banks pay higher interest rates on a loan with terms $\overline{t}$. However, the privilege of borrowing more frequently and perhaps larger sums may compensate these firms for the effective rate differential.[4]

Also, firms with large total assets frequently have many banking connections.[5] Individual banks may find their bargaining position slightly weakened by this greater competition, and consequently accept loans from larger firms, even though other features of their balance sheets are not as sound.

---

1. Robinson, op. cit., pp. 138-140.

2. Ibid., pp. 140-142.

3. Interpretation of the level of net worth as a measure of default risk is a bit treacherous. Net worth must be compared with the magnitude of misfortunes which a firm might have to face in order to be a realistic measure of default risk. The crude generalization in the text implicitly assumes that all firms may be expected to experience more or less similar unforeseen losses.

4. An excellent discussion of the role of deposit balances of borrowers may be found in a recent article by George Garvy. One implication of Garvy's argument is that banks desire balances in order to reduce their deposit instability, a factor which will be considered in Part 1 - 3. George Garvy, "Structural Aspects of Money Velocity," The Quarterly Journal of Economics, LXXIII (August 1959), pp. 443-446.

5. Cf. Katona, op. cit.

In Part 1 - 1, it was observed that banks try to make loans of moderate amounts, of short maturity, with high rates, and with collateral. Loan applicants also attempt to extract terms favorable to themselves. If the loan market is relatively competitive, banks and borrowers move each other to efficient frontiers. It was implied that the position of the bank's offer frontier is a function of characteristics of an applicant, summarized by $\underline{v}$, and characteristics of the bank, summarized by $\underline{u}$. If the frontier is not completely determined by $\underline{u}$, then some minimum credit standards exist for some point on the frontier.

The standards may be interpreted to be some minimum restrictions on some or all of the above variables. But obviously two borrowers need not possess identical balance sheets, income statements, etc., in order to be regarded as similar credit risks. Some trade-off among the above variables exists in the minds of credit analysts (banks). This trade-off is considered empirically in the subsequent sections.

1 - 3. Analysis of Bank Characteristics

Characteristics of banks also influence their willingness to grant loans. A bank isoquant of a loan offer function shows the minimum combination of characteristics which a bank must have before it is willing to make loans with terms $\underline{t}$ to borrowers with characteristics $\underline{v}$. It is a completely analogous concept to the borrower isoquant considered in Part 1 - 2.

The following characteristics of banks and their environment are believed to influence their willingness to lend: a bank's deposit level and stability, equity, and growth in deposits; the proportion of a bank's portfolio in loans and the distribution of its loans among industries; the maturity structure of its security holdings; the aggressiveness and specialization of a bank's lending officers; the interest rates on competing assets; the legal restrictions on its actions; the demand for a bank's loans; and the structure of its competition.

Banks are less willing to lend when their deposits are unstable. This behavior derives from banks' aversion to discounting, criticism by bank examiners, and emergency reserve adjustments. The probability that a given excess reserve position will be insufficient to meet weekly fluctuations is an increasing function of deposit instability. Consequently, banks suffering from very unstable deposits prefer a more liquid portfolio. Further, the average maturity of loans of these banks tends to be shorter. [1]

---

1. In response to the question, "If you feel it is proper to make term loans, are any limitations suggested in your mind as to the desirable percentage of this type of loan as compared with the aggregate of all commercial loans within the bank?," one lending officer responded: "There is no rule of thumb as to what the desirable percentage should be. Obviously for a bank with no time deposits and very volatile or seasonal demand deposits, the percentage should approach zero." "Term Loans," Bulletin of the Robert Morris Associates, Supplement, July 1957, p. 7.

In the appendix it is demonstrated with relatively weak assumptions that the ratio of variance of deposits to the level of deposits is an increasing function of bank size. Thus large banks are expected to be less willing to lend than small banks, if banks are averse to a high deposit variance.[1] Bank aversion to high deposit variance implies that this variable is an argument in a bank's utility function. It is assumed that a bank's expected utility is a function of the rate of return on its assets, the standard deviation of rates of return, and the variance of its deposits. A second conclusion from the appendix is that growing or declining banks experience a smaller deposit variance than banks whose deposits are unchanging.

Because time deposits are more stable in the short run than demand deposits, banks which have larger proportions of time deposits are expected to have larger proportions of loans in their portfolios.

Bank examiners and regulations frequently specify that a bank must maintain a certain proportion of its portfolio in liquid assets. Also, a bank often is not permitted to lend sums in excess of 10% of its equity to any one borrower.[2]

Considerations of portfolio balance also influence the willingness of a bank to lend. It is not wise for banks to have 100% of their portfolios in loans nor would they be prudent to put all their loans in one industry. Diversification reduces default risks.[3] But the aggressiveness and specialization of a bank's lending officers modify the concept of portfolio balance. Unfortunately, none of these factors can be quantified.

If a bank is a monopoly, then it may be able to discriminate among loan applicants. In this case the loan offer function is not independent of the demand for loans. Because an analysis of the demand for loans is beyond the scope of this paper, the present theory is relevant only for banks which have close competitors.

Banks are less willing to lend at terms $\bar{t}$ when interest rates on competing assets rise. One reason is simply that banks as profit maximizing institutions prefer to substitute higher yielding bonds for loans

---

1. The argument is not relevant for very large banks who have access to the federal funds market. However, for other banks the analysis suggests that larger banks will have a smaller and more conservative loan component in their portfolios.

2. Indeed, as the following quotation suggests, banks may prefer not to lend even this maximum amount: "A bank should not take term loans from any customer in an amount so great that its legal lending limit or its appraisal of the borrower's credit worthiness will make it unable or unwilling to handle its proper share of the customer's short term credit needs." George S. Moore, "Term Loans and Interim Financing," Beckhart, ed., op. cit., p. 253.

3. Cf. Markowitz, op. cit., Chs. 2, 4.

with unchanged yields. Other reasons have been suggested by writers concerned with the "availability" of credit.

Wallich[1] and Musgrave[2] have suggested that when bond interest rates rise, banks' expectations about the ability of borrowers to repay loans are altered:

> First there is the familiar point that central bank action directed at a rate increase tends to be an indication that the boom has run its course.[3]

Roosa[4] and Scott[5] have suggested that changes in bond interest rates increase uncertainty in the minds of bankers about yields on assets. Because banks dislike uncertainty, they are less willing to lend in periods of changing rates.[6]

Roosa[7] and Musgrave[8] believe that banks prefer not to sell bonds which are selling below their acquisition price. This aversion[9] to recording capital losses means that banks attempt to rebuild their secondary reserves, if bonds become "frozen in." Consequently, banks' willingness to lend is reduced.

Finally, Samuelson[10] and Kareken[11] have observed that increases

---

1. Wallich, op. cit., p. 765.

2. Richard A. Musgrave, "Credit Controls, Interest Rates, and Management of the Public Debt," Income, Employment, and Public Policy: Essays in Honor of Alvin H. Hansen (New York: W.W. Norton and Company, 1948), pp. 221–254.

3. Ibid., pp. 227–228.

4. R. V. Roosa, "Interest Rates and the Central Bank," Money, Trade, and Economic Growth: In Honor of John Henry Williams (New York: MacMillan, 1951), pp. 270–295.

5. Ira O. Scott, Jr., "The Availability Doctrine: Theoretical Underpinnings," The Review of Economic Studies, XXV (October 1957), pp. 41–48.

6. "But the potency of such changes comes from the impact of 'uncertainty' upon markets dominated by sensitive investors." Roosa, op. cit., p. 284.

7. Roosa, op. cit., p. 290

8. Musgrave, op. cit., p. 228.

9. Cf. "Bank Reactions to Security Losses," Monthly Review of the Federal Reserve Bank of Kansas City, June 1960, pp. 9–16.

10. Paul A. Samuelson, "Statement," U.S. Congress, Joint Committee on the Economic Report, Monetary Policy and the Management of the Public Debt, 82nd. Congress, 2nd. Session, 1952, Patman Hearings, pp. 691–698.

11. J. Kareken, "Lenders' Preferences, Credit Rationing, and the Effectiveness of Monetary Policy," The Review of Economics and Statistics, XXXIX (August 1957), pp. 292–302.

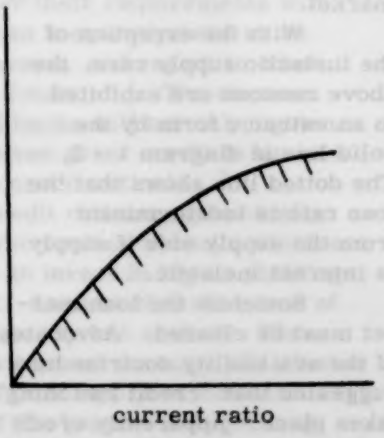
in bond interest rates temporarily reduce the market value of a bank's holdings of bonds. If banks had previously held a desired volume of secondary reserves (perhaps relative to their deposit variance), they may attempt to increase their security holdings when rate increases occur. This action reduces their willingness to lend.

Banks with different characteristics lend to borrowers summarized by $\overline{v}$ on terms $\overline{t}$. Bank isoquants are considered empirically in Sections 3 and 4.

1 - 4. Some Final Points

The endogenous and exogenous variables of the loan offer function have been introduced in Parts 1 - 1 through 1 - 3. It has been observed that banks substitute among endogenous variables and among each of the two classes of exogenous variables. Banks also substitute between these three groups of variables. For example, a bank is willing to alter one term of lending, say the amount of the loan, if a loan applicant can improve one of his balance sheet variables, perhaps his current ratio. Diagram 1 - 1 suggests that banks grant larger loans to borrowers with higher current ratios, if all other relevant variables are held constant. This curve is a complete analogue of the transformation curve in the theory of production.

Diagram 1 - 1



Similarly, substitution between sets $\underline{U}$ and $\underline{V}$ is easily imagined.

It is desirable to digress at this point and interpret the availability of credit doctrine with the loan offer function. In Part 1 - 3 it was observed that Wallich, Musgrave, Scott, Samuelson, Kareken, and Roosa were concerned with the impact of changes in government security interest rates on bank lending. By interpreting their arguments, it is possible to test this doctrine.

First, it should be observed that, with the exception of Samuelson, whose argument is dynamic, there is some doubt whether these authors are concerned with the rate of change in interest rates or simply the level of interest rates. I am not able to determine from their articles which is the variable that they believe causes changes in bank lending. If they are concerned with the rate of change of interest rates, then a dynamic argument must be made explicit in the loan offer function. Development of this argument is beyond the scope of the present paper. For this reason and because the subsequent empirical work is best directed towards analyzing the

impact of different interest rate levels, arguments of these authors are interpreted in terms of the impact on bank lending of different levels of competing asset (bond) interest rates.

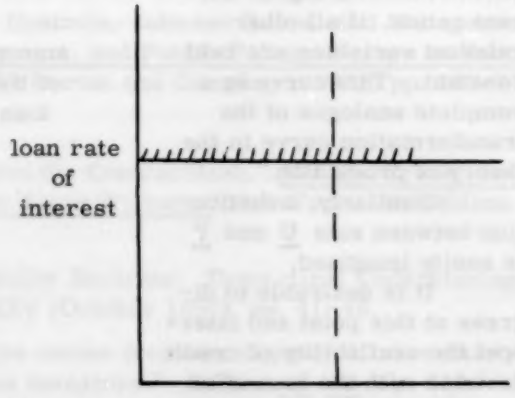
For various reasons, the above authors believe that the loan market is not cleared by loan interest rates alone, if interest rates on competing assets rise. Some believe that the supply schedule of loans is inelastic with respect to the loan rate of interest.[1] Others believe that loan interest rates are quite inflexible, and thus, although the interest elasticity of the supply schedule may not be negligible, it is not given a chance to work.[2] Usury ceilings may prevent interest rates from clearing markets. Finally, Professor Wallich has observed:

> At any time there is a more or less conventional range of rates on loans, negotiable and otherwise, which limits the maximum risk premiums that can be arranged for. Banks and investors usually prefer not to become involved in dubious situations, even if they believe their actuarial risk to be adequately compensated for.[3]

If banks associate dubious situations with high loan interest rates, regardless of rates on competing assets, again nonrate terms are clearing the market.

With the exception of the inelastic supply case, the above reasons are exhibited in an extreme form by the solid line in diagram 1 - 2. The dotted line shows that the loan rate is indeterminant from the supply side if supply is interest inelastic.

Diagram 1 - 2



Somehow the loan market must be cleared. Advocates of the availability doctrine have suggested that "credit rationing" takes place. Apparently credit rationing implies (1) that a higher percentage of loan applications are rejected and (2) that of those loans which are granted to borrowers characterized by $\overline{\underline{v}}$, the amounts are smaller, the maturities are shorter, and the security requirements are greater. The notion of credit rationing can be seen in the following quotations:

---

1. Ira O. Scott, Jr., "The Availability Doctrine," The Canadian Journal of Economics and Political Science, XXIII (November 1957), p. 536.

2. Kareken, op. cit., p. 295.

3. Wallich, op. cit., p. 768

> . . . the allocation of funds is guided by the banker's desire to allocate quotas to different uses of funds so as to obtain the right combination of risk and liquidity in his asset portfolio. . . . When a bank's lending capacity is curtailed, the quota available for each group of borrowers will be cut somewhat, but the quota for risk loans will tend to fall more sharply than that for credit to prime borrowers.[1]
>
> A change in interest rates may make itself felt less by affecting profit calculations, and through them the demand for funds than by affecting the behavior of financial institutions which lend the funds or act as intermediaries between the borrowers and the ultimate lenders. . . . As far as commercial banks are concerned a substantial drop in the official discount rate (accompanied by an increase in the liquidity of banks and a fall in the rates which they charge customers) may induce them to 'comb the market' more thoroughly than before. . . . Thus they may be more ready than before to lend to entrepreneurs who wish temporarily to finance the installation of new equipment by short-term credits with the intention of funding the credits later through security issues, or of repaying them out of working profits. . . . They may lower their requirements with respect to the 'current ratio' which is so frequently used as an index of the credit-worthiness of customers, and resort to other devices for creating an outlet for their funds. . . . The reverse applies more forcefully. . .commercial banks will not only tighten the conditions under which they themselves grant credit to business, but they will also often resort to credit rationing.[2]

Diagram 1 - 3 attempts to represent credit rationing graphically. Let there be a borrower characterized by $\bar{v}$ who applies to banks which are identical, except that some have opportunities to invest in higher yielding securities (which are assumed to be riskless). Assume that all terms of lending but the rate of interest and the loan maturity are settled. There are two cases: (1) loan interest rates are flexible and the supply is elastic and (2) loan interest rates are inflexible and/or the supply is inelastic. In case 1, nothing can be said about the ability of the borrower to get long term credits from the banks which can invest in higher yielding bonds. This is because, at a sufficiently high level of loan interest rates, applicants will be able to compensate these banks for the inconvenience of long term loans. The curve in diagram 1 - 3 will shift out to the dotted lines when loan interest rates rise.

In case 2, however, changes in loan interest rates can be ignored. The solid curve in diagram 1 - 3 then implies that, when competing rates

---

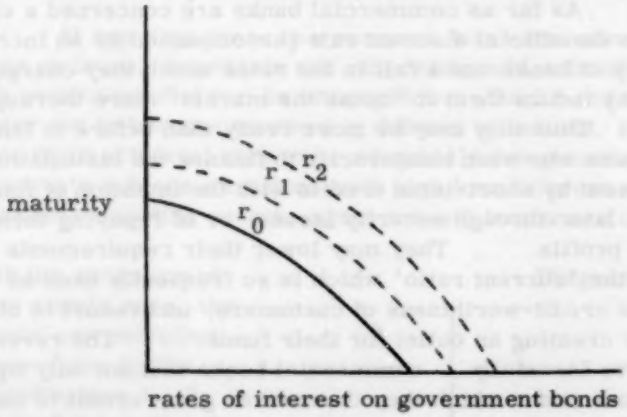
1. Musgrave, op. cit., p. 230.

2. F. A. Lutz, "The Interest Rate and Investment in a Dynamic Economy," The American Economic Review, XXXV (December 1945), pp. 828-829.

are high, loan applicants are unable to get certain long maturity loans.[1] Formulating lending behavior in terms of a loan offer function permits an empirical test of whether terms of lending do exhibit this hypothesized response to high rates of interest on government securities.

The form of the loan offer function must be made explicit if the theory is to be subjected to empirical test, so we return to the main argument.

Diagram 1 - 3



Let:

R = the loan rate of interest (in percent).
M = the maturity of the loan (in months).
A = the amount of the loan (in dollars).
S = 1 if the loan is secured, 0 otherwise.
W_i = the ith relevant characteristic of loan applicants, $i = 1, 2, \ldots, I$.
Z_j = the jth relevant characteristic of lending banks, $j = 1, 2, \ldots, J$.

If no other terms of lending are relevant to a bank, equation 1 - 1 may be written as:

1 - 2) $F(R, M, A, S) = G(W_1, W_2, \ldots, W_I; Z_1, Z_2, \ldots, Z_J)$.

Can F and G be expressed simply? It is not likely that they are linear in the variables. In the case of G, a firm could achieve an infinite current ratio by paying off its current liabilities, but banks do not lend large sums of money for extended periods simply because of this transaction. Similarly, firms with very large net earnings are not able to obtain loans at

1. To place Musgrave properly in this interpretation, one must assume that longer maturity loans are more risky.

arbitrarily low rates of interest. The fact that a firm regularly earns one million dollars greatly influences a bank's appraisal of its credit worthiness, but the bank will not be as greatly impressed by the difference between regular earnings of ten and eleven million dollars. Similar arguments can be constructed for F.

For these reasons non-linear (polynomial or logarithmic) formulations of F and G are more appealing.[1] If F and G are polynomial, the theory should specify the order of the polynomial and the character of the interactions between the various arguments of the function. Some broad conjectures about the relationships can be made.

Equations expressing each of the four terms of lending as functions of Z's and W's, ignoring other terms of lending, were estimated for a subsample of loans with multiple regression techniques. Various functional forms were attempted. Inspection of the resulting residuals suggested that they were distributed lognormally. Throughout the remainder of this paper, it is therefore assumed that F and G are logarithmic in their arguments.[2] The theory of the loan offer function does not preclude F and G having logarithmic forms.

In order to determine whether equation 1 - 1 or the regressions mentioned in the preceding paragraph (hereafter called partial loan offer functions) are identifiable, it is necessary to consider the demand for terms of lending. By an argument quite analogous to the theory of bank lending, loan applicants are believed to have desires for certain rates of interest, maturities, amounts, security requirements, etc. Each borrower possesses some function, a loan asking function, which characterizes his desires in terms of his needs. The functions differ substantially among applicants. By an argument presented by E. J. Working,[3] identification is possible if the variance of the stochastic terms in the supply function (loan offer function) is substantially smaller than the corresponding variance in the demand function (loan asking function). Identification is easier if

---

1. These non-linear forms are not totally satisfactory in the case of the loan rate of interest. Because of the existence of a prime rate, a kink in the offer frontier will occur when borrowers are sufficiently powerful to begin to borrow at prime. Empirically, this problem may be avoided by using probit regression techniques of estimation. If this elaboration is ignored, the misspecification is probably not very serious. Cf. James Tobin, "Estimation of Relationships for Limited Dependent Variables," Econometrica, XXVI (January 1958), pp. 24-36.

2. It is necessary to redefine S if logarithms of that variable are to be taken. If S = 1, it was set equal to 10; if S = 0, it was set equal to 1.

3. E. J. Working, "What Do Statistical 'Demand Curves' Show?" Readings in Price Theory (Chicago: Richard D. Irwin, Inc., 1952), pp. 97-118.

different borrowers have different asking functions.

What basis can there be for assuming that the stochastic terms in the asking function (or partial asking functions) are much larger than those in the offer function (or partial offer functions)? First, a borrower desires loans for reasons which are related to seasonal or cyclical fluctuations in his industry. Second, treasurers of different firms are requesting loans and there is no compelling reason for these individuals to behave similarly. Finally, many of the factors with which banks are concerned (an applicant's credit position or a bank's deposit instability) are unrelated to a firm's desire for terms of lending.

A given bank, however, forces its lending officers to follow certain common techniques of credit analysis. Greater experience with processing loan applications on the part of these officers results in precision (smaller stochastic variance) in lending. Finally all commercial banks are subjected to more similar external forces than are a sample of borrowers drawn from a wide range of industries. It is therefore possible to ignore demand or the loan asking function.

A loan offer function may be estimated directly by the method of canonical correlation. This technique involves defining two canonical variates:

$$q_1 = (\text{Log } R)' + k_1(\text{Log } M)' + k_2(\text{Log } A)' + k_3(\text{Log } S)' + u.$$ [1]

$$q_2 = \sum_{i=1}^{I} a_i(\text{Log } W_i)' + \sum_{j=1}^{J} a_{I+j}(\text{Log } Z_j)' + v.$$

The stochastic variable u is assumed to be independent of the terms of lending, but not independent of characteristics of the borrower and lending bank. Similarly, v is assumed to be independent of those characteristics, but not independent of the terms of lending.

The computational procedure yields those a's and k's which maximize the correlation between q_1 and q_2 when q_1 and q_2 are normalized to have unit variance.[2] The results of applying this technique to five samples of loans are reported in Section 4.

Unfortunately, it is not possible to obtain standard errors for the a's and k's. But some basis is necessary for testing which characteristics of banks and applicants are relevant in lending decisions. A test, which is by no means ideal, is whether characteristics of banks and borrowers are significantly related to <u>each</u> of the terms of lending. In Sections 2 and 3 partial loan offer functions are estimated from five samples of loans in order to make this test.

---

1. Primes (') on variables indicate that they are defined as deviations from their means.

2. Harold Hotelling, "Relations Between Two Sets of Variates," <u>Biometrica</u>, XXVIII (December 1936), pp. 321-377.

Finally, if there is no substitution between terms of lending, then the partial loan offer functions completely characterize bank lending behavior. The only difference between the partial loan offer function and the loan offer function is that the latter permits substitution among terms of lending.

2. Analysis of Term Lending

This section presents statistical results of an analysis of commercial bank term lending. Data were collected from a small number of large banks on individual term loans granted during the period January 1955 through October 1957.[1] The data include information about terms of lending, the credit position of the loan applicant, and previous relationships between the bank and the applicant.[2]

2 - 1. Properties of Term Loans

A term loan is a loan transaction which has a final maturity exceeding one year, which may or may not be secured and which usually specifies repayment in monthly, quarterly, semi-annual, or annual payments. Moreover, it is evidenced by a note or notes embodying the terms and conditions of the loan, or is subject to a separate formal agreement between the borrower and the lender.[3]

Term lending became popular among bankers in the United States during the depression when demand for loans was low. Because of the sharp break with traditional "seasonal" financing, an extensive discussion of this method of lending evolved in the contemporary banking literature and has continued to the present day. Bankers have been cautious in granting term loans and have kept detailed statistical records of their experience with them.

Current interest in term lending and superior statistics about past term lending are compelling reasons for studying this class of loans.[4]

---

1. I am most indebted to the co-operating banks for their very considerable advice and labor while making these data available. Their overworked credit departments made substantial contributions to this research effort. Although it is inappropriate to reveal the names of the co-operating institutions, their interest in academic research is a tribute to themselves and the banking industry.

2. Information on sampling problems and techniques may be obtained by corresponding with the author.

3. American Bankers Association, Term Lending by Commercial Banks, Credit Policy Commission (New York: American Bankers Association, 1957), p. 5.

4. A. Hayes, "A Breathing Spell for Monetary Policy," Monthly Review of The Federal Reserve Bank of New York, XLII (June 1960), pp. 100-104.

Term loans have additional desirable properties. They are frequently large, which permits an investigator to ignore costs of lending. Terms of individual term loans vary substantially. Borrowers with considerably different credit positions are accorded term loan financing. Large commercial banks grant a sufficiently large number of term loans to permit within-bank analyses.

The quantitative significance of term lending at large banks in New York City may be seen in table 2 - 1.

Table 2 - 1

Business Loans at Large New York City Banks
on Selected Dates[1]
(in billions of dollars)

	10/5/55	10/16/57	10/14/59	10/12/60
All business loans	7.4	10.4	10.0	10.2
Term loans	3.5	5.2	5.7	5.6

1. Monthly Review of The Federal Reserve Bank of New York, XLIII (February 1961), p. 31.

Differences in term lending behavior among large commercial banks may be attributable to differences in bank preferences for term loans, differences in their portfolio positions at the time of lending, differences in bank deposit structure, etc. The relevance of results in this section to other banks depends upon the sensitivity of the estimated parameters to these omitted variables. Because the banks which are considered were not drawn in a random fashion, generalization of what follows to the population of banks is hazardous.

2 - 2. Hypotheses

Costs of data collection limited the volume of information about a borrower which could be obtained. In particular, no information was obtained about a borrower's net worth, inventories, liquid assets, or debt. Among the variables which were recorded are the following:

P_m = average losses of the applicant over the 3 to 5 years immediately preceding the loan (5 if available) if the applicant experienced net losses over that period, $P_m = 1$ otherwise. (in dollars per year)

P_p = average net profits of the applicant over the 3 to 5 years immediately preceding the loan (5 if available) if the

applicant experienced net profits over that period, $P_p = 1$ otherwise. (in dollars per year)

C = current ratio (ratio of current assets to current liabilities) of the applicant at the close of the applicant's fiscal year immediately preceding the loan.

E = average demand deposit balances of the applicant at the lending bank during the year before the loan, $E = 1$ if no balances existed. (in dollars)

Y = number of years which the applicant had continuously maintained a demand deposit relationship with the bank at the time of the loan, $Y = 1$ if there was no continuous relationship. (in years)

$\bar{R}$ = prime rate of interest at the time of the loan. (in percent)

W_m = negative of the average working capital of the applicant over the 3 to 5 years immediately preceding the loan (5 if available) if the applicant reported negative average working capital over that period, $W_m = 1$ otherwise. (in dollars)

W_p = average working capital of the applicant over the 3 to 5 years immediately preceding the loan (5 if available) if the applicant reported positive average working capital over that period, $W_p = 1$ otherwise. (in dollars)

$\left(\frac{dP}{dt}\right)_m$ = negative of the average annual change in the applicant's profits over the past 3 to 5 years (5 if available) if the change was negative, $\left(\frac{dP}{dt}\right)_m = 1$ otherwise. (in dollars per year per year)

$\left(\frac{dP}{dt}\right)_p$ = average annual change in the applicant's profits over the past 3 to 5 years (5 if available) if change was positive, $\left(\frac{dP}{dt}\right)_p = 1$ otherwise. (in dollars per year per year)

$\left(\frac{dW}{dt}\right)_m$ = negative of the average annual change in the applicant's working capital over the past 3 to 5 years (5 if available) if change was negative, $\left(\frac{dW}{dt}\right)_m = 1$ otherwise. (in dollars per year)

$\left(\frac{dW}{dt}\right)_p$ = average annual change in the applicant's working capital over the past 3 to 5 years (5 if available) if change was positive, $\left(\frac{dW}{dt}\right)_p = 1$ otherwise. (in dollars per year)

H = total assets of the applicant at the close of the fiscal year immediately preceding the loan. (in dollars)

V_p = variance of the applicant's annual profits over the 5 years immediately preceding the loan. (in dollars squared per year squared)

A_o = the amount of loans obtained from other banks (their participation) in the term loan transaction, if more than one bank is involved. The variable is not defined for

loans involving only one bank. (in dollars)

E_a = amount of deposit balances carried by some affiliates of the applicant with the lending bank. (in dollars)

It is necessary to define two variables to represent mean profits, mean working capital, and trends in profits and working capital because of the decision to use logarithms reported in Section 1. Similarly, null values of balances, years of relationship, etc., must be redefined. This is because logarithms of non-positive numbers are undefined.[1]

The following hypotheses may be inferred from the discussion in Section 1. Subsequently the hypotheses are defended in terms of the rate of interest. It is hypothesized that each of the following cause the rate of interest on the loan to increase, and/or the maturity of the loan to decrease, and/or the amount of the loan to decrease, and/or likelihood of security to increase.[2]

1) Ceteris paribus, an increase in an applicant's average losses P_m or a decrease in his average profits P_p.
2) Ceteris paribus, a decrease in an applicant's current ratio C.
3) Ceteris paribus, a decrease in an applicant's average deposit balance E.
4) Ceteris paribus, a decrease in the length of time Y over which the applicant has continuously maintained deposits with the bank.
5) Ceteris paribus, an increase in the prime rate $\bar{R}$.
6) Ceteris paribus, an increase in an applicant's average negative working capital W_m or a decrease in his average positive working capital W_p.

---

1. The procedure is definitely not ideal. There is no reason to believe that bankers regard zero as a particularly significant number. It may be that loans to firms with profits less than ten thousand dollars have a different profit elasticity than loans to firms with profits above ten thousand dollars. Then logs should be taken of profits minus ten thousand. Although this problem is generally intractable, C. I. Bliss verbally suggested an iterative scheme for those cases where "log zero values" are below the minimum profit figure contained in the sample. His method was developed for estimating α and β in the following equation: Log $y = \alpha$ Log $(x - \beta) + u$. Unfortunately, the multivariate extension of this technique did not converge and it became necessary to adopt the above convention.

2. Because S is a dummy variable, statements involving S require some interpretation. If regression equations predict a value of S close to unity, it is believed that banks are more likely to insist upon security. Interpretations about the variance of residuals of regressions involving S are also ambiguous. Throughout this paper it is assumed that regressions involving S can be analyzed as if S is a continuous variable.

7) Ceteris paribus, an increase in an applicant's average annual decline in profits $\left(\frac{dP}{dt}\right)_m$ or a decrease in his average annual growth in profits $\left(\frac{dP}{dt}\right)_p$.

8) Ceteris paribus, an increase in an applicant's average annual decline in working capital $\left(\frac{dW}{dt}\right)_m$ or a decrease in his average annual growth in working capital $\left(\frac{dW}{dt}\right)_p$.

9) Ceteris paribus, a decrease in an applicant's total assets H.

10) Ceteris paribus, an increase in the ratio of losses to assets P_m/H or a decrease in the ratio of profits to assets P_p/H.

11) Ceteris paribus, an increase in an applicant's borrowings from other banks (participations) A_o.

12) Ceteris paribus, an increase in the variance of an applicant's profits V_p.

13) Ceteris paribus, a decrease in balances of affiliates of the applicant E_a.

Hypothesis 1 states that if there are two firms, identical in every essential respect except that one earned larger after tax profits in recent years, applying to a bank for a loan, then the firm with larger profits will be charged a lower rate of interest. The rationale for hypothesis 1 was presented in Section 1. Average profit figures of the applicant were employed rather than profits in the most recent year in order to avoid undue noise in the equations and because the credit literature[1] advocates the use of average profit figures.[2]

Hypothesis 2 states that if two firms, identical in every essential respect except that the current ratio differs between the firms, apply to a bank for a loan, then the firm with the larger current ratio will be charged a lower rate of interest. The current ratio is particularly significant because current liabilities may be viewed as a measure of expected short term future outpayments which a firm will have to make. The ratio of current assets to current liabilities is therefore a measure of the extent to which these outpayments can be covered without recourse to outside borrowings.

The current ratio and other variables to be considered in this section should be adjusted for the character of an applicant's business. Thus

---

1. American Bankers Association, op. cit., pp. 9-10.

2. The fact that continuous profit figures were reported for only three years in some cases may bias the estimated profit elasticities. The bias is believed to be negligible.

firms with a large current ratio may be more prone to bankruptcy than firms with a smaller current ratio, simply because the current liabilities of the former are more volatile. Some adjustment for interindustry differences is made by banks which use techniques advocated by the Robert Morris Associates.[1] These techniques involve comparing a firm's current ratio with the average current ratio of similar firms in the same industry. No attempt was made to utilize such techniques in this section because industrial classifications are ambiguous and the typical term loan borrower had few comparable firms within his "industry."

A second inadequacy of the current ratio (and also working capital) is that current assets include inventories. If a firm's inventories are large, his current ratio may be large, but he may be close to bankruptcy. If the hypothesis is to have the indicated sign, it is necessary to assume that inventories are controlled. A more detailed analysis would include inventories as a separate variable.

Hypothesis 3 states that if two firms, possessing identical income statements and balance sheets and differing only in that one firm maintains a larger demand deposit balance with the lending bank than the other, apply to a bank for a loan, then the bank will charge a lower rate to the holder of the larger balance. One reason for this behavior is that firms carrying larger deposit balances are expected to do so in the future and thus in order to charge both applicants the same effective rate a bank must act in accordance with the hypothesis. A more interesting argument is associated with the fact that applicants who keep idle balances with a bank perform a favor for the bank by reducing its deposit instability. Banks may prefer to charge such firms a slightly lower rate rather than risk the loss of their accounts.

Measurement difficulties obscure the significance of E, the applicant's previous year average balance. Information about activity of balances could not be obtained. If a great number of checks were drawn on the account, bookkeeping expenses and excessive deposit instability may negate the desirability of a large balance. This qualification is not likely to be too serious because bankers impose service charges after a certain level of activity.

A second problem arises because it was not possible to obtain statistics about an applicant's current or previous short term borrowings. The argument for hypothesis 3 is concerned with an applicant's deposit balance net of other borrowings. Use of E therefore unavoidably involves a misspecification of the loan offer function. Despite these objections, it is important to note that banks keep records of gross deposit balances rather than net deposit balances and from comments by bankers I infer that E is a reasonable proxy for net balances.

---

1. Cf. Robert Morris Associates, 1958 Statement Studies (Philadelphia: Robert Morris Associates, 1959).

A final problem relates to the complexity of organization of contemporary borrowers. A corporation may have five or six affiliated concerns, each of which carry deposit balances with the bank. Hypothesis 13 represents a crude attempt to take this factor into account. In the case of certain loans, the banks voluntarily reported "conspicuous" affiliated balances which were factors in the bank's decision to lend. Needless to say, reporting of such affiliated balances represented an arbitrary judgment. It is quite difficult to associate balances of a huge complex with any particular firm. Empirical examination of affiliated balances is at best hazardous.

Hypothesis 4 states that if two firms, possessing identical income statements, balance sheets, and being similar in every other essential respect except that one had maintained a deposit account with the bank for a longer period, apply to a bank for a loan, then the bank will charge a lower rate of interest to the applicant with the longer account relationship. The reasoning is that banks have more knowledge of old customers and their businesses. Greater knowledge reduces uncertainty associated with a loan. Banks ordinarily compensate for uncertainty by charging higher interest rates.

It is of course true that merely carrying a deposit with a bank for an extended period need not increase a bank's knowledge of an applicant. Previous loans to the applicant, however, would have conveyed information to the bank. It is assumed that old depositors are more likely to have borrowed in the past. [1]

Hypothesis 5 states that if two identical firms apply to a bank for a loan at two different points in time at which the bank is in the same position except that the prime rate of interest has risen, then the bank will charge a higher rate when prime is higher. This is because banks regard prime as a floor for loan interest rates and tend to set rates at prime plus some differential for more risky borrowers. Because prime is essentially a convention among large banks, the hypothesis does not apply to smaller banks.

The reader may object that movements in the prime rate have little relevance to the amount or maturity of a loan. Large banks move prime relatively frequently as can be seen in diagram 2 - 1. In interviews, however, lending officers in commercial banks concede that interest rates above prime are somewhat less flexible. If banks are initially in equilibrium and higher loan rates are less flexible when prime rises, banks may be expected to ration credit[2] with respect to other terms of lending in order to clear the loan market.

---

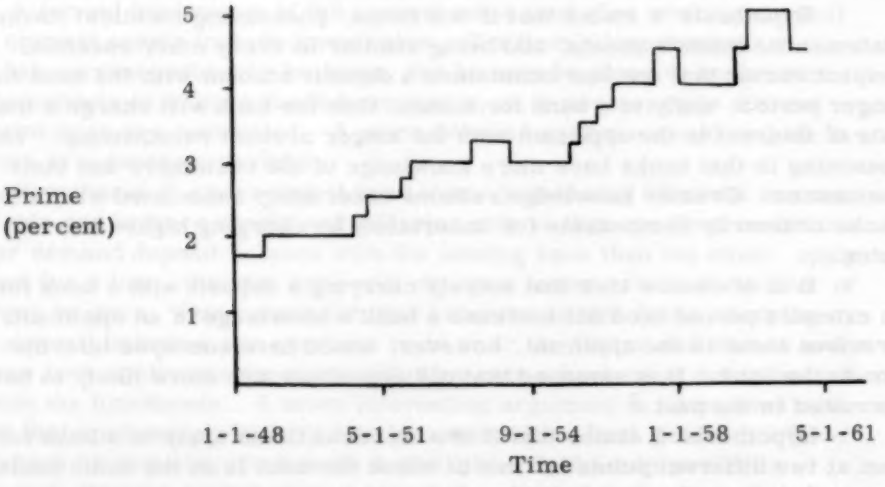
1. An alternative hypothesis implies an opposite relationship between the years of deposit relationship and terms of lending. This would follow if banks compete for new accounts by offering low rates of interest. Old depositors, due either to inertia or unwillingness to change banks because of other bank services, are charged slightly higher rates. A priori, this explanation does not seem likely and the conjecture is not tested.

2. This argument is consistent with the availability of credit doctrine.

A second argument which does not require inflexibility of certain rates is that banks prefer to substitute less risky loans at higher levels of market interest rates. Empirically, it is not possible to discriminate between these explanations.

Hypothesis 6 states that if two firms, with identical income statements, etc., and with balance sheets which are identical in every essential

Diagram 2 - 1[1]



respect except that one has larger working capital (current assets minus current liabilities), apply to a bank for a loan, then the bank will charge a lower rate of interest to the firm with larger working capital. This hypothesis is quite similar to hypothesis 2 and is an alternative way of formulating the current position of a firm. Net working capital differs from the current ratio in that it does not show the relative magnitudes of current assets and current liabilities.

Hypothesis 7 states that if two firms, identical in every essential respect except that one has been experiencing a larger average annual growth in net profits or a smaller average annual shrinkage in net profits, apply to a bank for a loan, then the bank will charge a lower rate of interest to the firm whose profits have been growing most or declining least. Firms with growing profits are excellent prospects for future balances and loan business. Banks may interpret growth in profits as evidence of superior management. A growing stream of profits, which is expected to continue growing, provides an increasing safety margin

1. Data necessary to construct diagram 2 - 1 were secured from a confidential inter-office memorandum in a large commercial bank. A similar chart appeared in The Wall Street Journal, August 29, 1960.

against unanticipated setbacks in a firm's business.

The variables $\left(\frac{dP}{dt}\right)_m$ and $\left(\frac{dP}{dt}\right)_p$ were obtained by regressing each borrower's profits on time.[1] The resulting slopes are the variables.

Hypothesis 8 is perfectly analogous to hypothesis 7 except that it refers to annual changes in working capital rather than annual changes in profits. The variables $\left(\frac{dW}{dt}\right)_m$ and $\left(\frac{dW}{dt}\right)_p$ were again obtained as slopes of regressions. Growth in working capital may result from growth in inventories, payment of accounts payable, collection of accounts receivable or from growth in cash assets attributable to profit flows, depreciation flows and reductions in investment, dividends, etc. If working capital is improving over time and is expected to continue to grow, again an increasing safety margin against unanticipated setbacks in the firm's business exists.

Hypothesis 9 states that if two firms, identical in every essential respect except that one possesses larger total assets, apply to a bank for a loan, then the bank will charge a lower rate to the firm with larger total assets. The reason for this behavior is that large firms have more banking connections and thus are in a superior bargaining position.[2]

Hypothesis 10 argues that if two firms, identical in every essential respect except that one possesses a higher ratio of average profits to assets or a lower ratio of average losses to assets, apply to a bank for a loan, then the bank will charge a lower rate to the applicant possessing the greater profit rate. This hypothesis differs from hypothesis 1 in that it asserts that banks are concerned with the "efficiency" with which firms utilize their assets rather than with dollar flows. In future research, a more meaningful profit rate might be expressed as the ratio of profits to net worth.

Hypothesis 11 states that if two firms, identical in every essential respect except that one is simultaneously obtaining loans from other banks, apply to a bank for a loan, then the bank will charge a lower rate to the firm which is not simultaneously borrowing from other banks. The reason is that simultaneous borrowings increase the ratio of debt to equity of the applicant thereby increasing the probability of default.

Hypothesis 12 states that if two firms, identical in every essential respect except that one experiences greater variance of profits over time, apply to a bank for a loan, then the bank will charge a lower rate of interest to the firm with the lower profit variance. The reason is that

---

1. The equation of the regression is $P = a + bt + \epsilon$.

2. Total assets is correlated with many other balance sheet variables which have been omitted from this analysis. Assets might be viewed as a proxy for net worth, debt, current assets, etc. Unfortunately, these variables do not affect a bank's willingness to lend in the same way and thus the sign of the relationship becomes ambiguous.

future instability of earnings may imply increased probability of bankruptcy and it may cause a delay in loan repayment (not necessarily associated with bankruptcy). This variable is the residual variance about the regression line from which $\left(\frac{dP}{dt}\right)_m$ and $\left(\frac{dP}{dt}\right)_p$ were obtained. It was only computed for those observations containing five consecutive annual profit figures. In interviews, bankers reported that variability of profits was an undesirable feature of loan applicants, but were unable to specify what measure of variability they employed. Variance of yearly profits is assumed to be an appropriate measure.

Hypothesis 13 states that if two firms, identical in every essential respect, both having affiliates or subsidiaries which carry balances with a bank, apply to that bank for a loan, then the firm with larger affiliated balances will be charged a lower rate of interest. The reason for this hypothesis was presented while discussing hypothesis 3.

2 - 3. Empirical Results

Partial loan offer functions were estimated from data on term loans for each of three large commercial banks. Term loan information collected from smaller commercial banks was either incomplete or of insufficient quantity to permit analysis. Hypotheses 1 through 10 were tested by examining coefficients of estimated partial loan offer functions. Hypotheses 11 through 13 were tested by examining residuals of the estimated partial offer functions.[1]

Use of partial loan offer functions for tests of the above hypotheses is not without objection. A variable may be insignificant in each of the partial offer functions and yet be significantly related to some linear combinations of terms of lending. It is arbitrarily assumed that, if any hypothesis is significant in more than one of the estimated partial offer functions, the hypothesis is accepted.[2] Tests of coefficients are one tailed

---

1. A_o, E_a, and V_p were reported for a smaller number of loans than were the other variables. It was assumed that coefficients associated with these other variables in each of the four partial offer functions were independent of the reporting and magnitude of A_o, E_a, and V_p. The assumption permits estimation of these other coefficients with all observations.

2. A decision rule is necessary because the banks were not selected randomly. With twelve partial offer functions, the probability of a random variable being significant at the .025 level at least once is approximately .26. The probability of a random variable being significant in two or more regressions is approximately .035.

The reader might question why a simple pooling of "t" ratios would not be a more elegant test than the procedure reported in the text. The difficulty with pooling is that a variable may only (1) be strongly

tests and a coefficient is judged significantly greater or less than zero if the probability of the event occurring when the null hypothesis is true is less than .025.

Table 2 - 2 reports regressions of the loan rate of interest on profit and profit rate variables, the current ratio, applicant's balances, years as a depositor, and the prime rate for each of the three banks. Tables 2 - 3 through 2 - 5 contain the corresponding regressions of maturity, amount, and the security dummy variable.

All regressions in the following tables are significant at the .05 level. The number of observations ranges from 60 to 130.[1]

Hypothesis 1 is accepted because the profit variable P_p is significant in eight of twelve regressions and the loss variable P_m is significant in one. Unfortunately, had the hypothesis about the loss variable been of the opposite sign, that variable would have been significant in five of the regressions. In other words, the results suggest that firms which experience large losses or profits are lent money on easier terms than firms with small net profits. The loss paradox apparently derives from misspecification.

One explanation is that profits are highly correlated with scale variables, total assets, and net worth. Doubt is cast on this explanation by the credit literature[2] on term loans which emphasizes earnings rather than net worth. When total assets is added to the above regressions, the fit is not improved. Nevertheless, the P_m coefficients might change sign if the right scale variable were employed.

A second explanation which seems more plausible derives from the observation that firms which lose substantial sums of money for periods of three to five years are exceptional firms. They are even more exceptional when commercial banks will lend them money for a period of years. Their exceptional characteristics may include development of new products, particularly capable management, etc. Had these

---

related to one term of lending or (2) be significant in the case of one bank. It is quite possible that such variables would not prove significant when pooled. It is my view that these two cases should not be ignored. The alternative is to consider each of the three banks separately, a procedure which would unduly extend this paper.

1. The reader is cautioned that the sample of loans used to determine the functional form is included in regressions of bank 1. Approximately fifty percent of the observations from bank 1 were used to determine the appropriate functional form. A purist would insist that these observations not be included in subsequent tests. In fact, regressions estimated from the fifty percent which were previously used and from the other half of the sample were not very different.

2. American Bankers Association, loc. cit.

TABLE 2 - 2[1]

Regressions of Logarithm of Loan Rate of Interest

	Bank 1	Bank 2	Bank 3
R^2	.768	.801	.740
Constant	.29799	.22819	.50641
	(.03879)	(.07617)	(.05441)
Log P_m	-.021376	.0077583	-.048994
	(.010546)	(.0126343)	(.02144)
Log P_p	-.028141 *	-.026016 *	-.051199 *
	(.003764)	(.007273)	(.005624)
Log (P_m/H)	.013766	.17353 *	.044202
	(.035608)	(.06243)	(.039758)
Log (P_p/H)	-.0040334	.0000387	.019266
	(.0120053)	(.0192769)	(.012640)
Log C	-.0035993	.0050930	.027917
	(.0094913)	(.0152388)	(.011215)
Log E	-.0022438	-.0012317	.0057964
	(.0022897)	(.0035640)	(.0038054)
Log Y	.0013883	.011829	-.041044 *
	(.0081666)	(.012885)	(.013261)
Log $\bar{R}$	.88445 *	.94109 *	.82513 *
	(.04975)	(.07117)	(.07837)

1. Numbers in parentheses are the standard errors of the coefficients in this and subsequent tables.

* implies significance at .025 level.

characteristics been included in the regressions, perhaps P_m would have coefficients of a different sign.

Concentrating on the P_p coefficients, the estimated profit elasticity of the loan rate of interest varies between .026 and .051 and differs significantly between the banks. Thus a tenfold increase in a firm's profits will permit it to borrow term money at a rate of interest which is lower by approximately one fourth to one half of a point.

TABLE 2 - 3

Regressions of Logarithm of Loan Maturity

	Bank 1	Bank 2	Bank 3
R^2	.300	.507	.252
Constant	2.7423	2.4538	1.9245
	(.2805)	(.4948)	(.3262)
Log P_m	.022349	.077759	.089841
	(.076251)	(.082036)	(.128551)
Log P_p	-.01642	.097984 *	.061644
	(.2722)	(.047225)	(.033717)
Log (P_m/H)	.087133	-.39020	.033918
	(.257463)	(.40538)	(.238343)
Log (P_p/H)	.13893	-.0091561	-.069556
	(.08780)	(.1251702)	(.075777)
Log C	.25379 *	.23589 *	.21279 *
	(.06863)	(.09895)	(.06724)
Log E	.0099599	-.027579	-.014139
	(.0165557)	(.023142)	(.022814)
Log Y	-.001404	.10170	.019700
	(.0590489)	(.08367)	(.079500)
Log $\overline{R}$	-1.6371 *	-2.4696 *	-1.1492 *
	(.3597)	(.46213)	(.4698)

* implies significance at .025 level.

TABLE 2 - 4

Regressions of Logarithm of Amount of Loan

	Bank 1	Bank 2	Bank 3
R^2	.436	.536	.479
Constant	3.8782	3.4590	2.7861
	(.3947)	(.6300)	(.6623)
Log P_m	.34318	.44928	.62381
	(.10730)	(.10451)	(.26101)
Log P_p	.30118 *	.34179 *	.44388 *
	(.03830)	(.06016)	(.06846)
Log (P_m/H)	-.23291	.31845	-.060286
	(.36231)	(.51641)	(.483947)
Log (P_p/H)	-.22484	-.0209343	-.29203
	(.12216)	(.159452)	(.15386)
Log C	-.11031	-.074821	-.17331
	(.09658)	(.1260521)	(.13652)
Log E	.071128 *	.064824 *	.052576
	(.023298)	(.029481)	(.046322)
Log Y	-.082879	.0024384	-.059843
	(.083096)	(.1065818)	(.161421)
Log $\bar{R}$	-.24783	.40405	-.22943
	(.50619)	(.58871)	(.95400)

* implies significance at .025 level.

In the case of bank 2, increases in a firm's profits induce that bank to lend at longer maturities. A similar statement does not characterize the other banks.

All three banks grant larger loans when an applicant's profits rise. Profit elasticities with respect to amount vary from .30 to .44 and the

coefficients differ significantly between the banks.

In the case of bank 1, increases in a firm's profits and decreases in a firm's losses are associated with increased likelihood of security. Similar statements do not apply to the other banks.

Hypothesis 2 is accepted because coefficients of C are significant in

TABLE 2 - 5

Regressions of Logarithm
of Security "Dummy" Variable

	Bank 1	Bank 2	Bank 3
R^2	.238	.264	.222
Constant	.64815 (.42578)	-.14593 (.82473)	.64072 (.61148)
Log P_m	.31561 * (.11575)	.20416 (.13680)	.23161 (.24098)
Log P_p	-.097139 * (.041314)	.0032402 (.0787505)	-.071817 (.063206)
Log (P_m/H)	1.0529 * (.3908)	-.15262 (.67600)	.43284 (.44680)
Log (P_p/H)	-.24308 (.13177)	-.55109 * (.20873)	-.23919 (.14205)
Log C	-.10298 (.10418)	.10804 (.16500)	-.20391 (.12604)
Log E	.0050158 (.0251320)	.0076621 (.0385914)	.076944 (.042767)
Log Y	-.12322 (.08964)	-.10592 (.13952)	-.31464 * (.14903)
Log $\bar{R}$	-.055012 (.54604)	-.79358 (.77063)	-.51984 (.88077)

* implies significance at .025 level.

each bank's maturity regression. Banks lend further into the future as an applicant's current ratio improves. The current ratio elasticity of maturity varies from .21 to .25, a difference well within the standard errors of each of the estimates. The coefficient of C has the wrong sign in the rate regression of bank 3. There is no obvious reason for this and it is assumed to be spurious.

Hypothesis 3 is accepted because coefficients of E are significant in the amount regressions for banks 1 and 2. Firms holding larger balances at these banks tend to be granted larger loans. Balance elasticities in both cases are approximately .07.

Hypothesis 4 is accepted although significant coefficients of Y appear only in regressions involving bank 3. Old depositors are able to secure loans at lower rates of interest and more frequently without security.

Hypothesis 5 is accepted. Coefficients of $\bar{R}$ are significant in the interest rate and maturity regressions for all three banks. The prime rate elasticities of the loan rate of interest are .88, .94, and .83 for banks 1, 2, and 3, respectively. The lower two elasticities are significantly different from unity. This should not necessarily be interpreted as evidence of credit rationing. Bankers frequently express loan rates to borrowers as prime, prime plus a quarter, prime plus a point, etc. The logarithmic formulation assumes that differentials from prime increase proportionately with the level of prime. Consequently, if many loans are made at rates above prime and the differentials are independent of prime, the coefficient of $\bar{R}$ would be expected to be less than unity.

Coefficients of $\bar{R}$ in the case of the maturity regressions are revealing. Increases in the prime rate tend to shorten the maturity of new loans.[1] The result suggests that loan interest rates did not or could not rise sufficiently above prime to induce banks to grant very long maturity term loans. Because movements in the prime rate are correlated with movements in bill rates, discount rates, bond rates, etc.; one implication of restrictive credit policies by monetary authorities may be a reduction in the willingness of banks to grant longer term loans. This result might also derive from the Wallich-Musgrave interpretation of the availability doctrine if it is believed business expectations are inversely related to the level of interest rates. Other coefficients of $\bar{R}$ are not significantly different from zero.[2]

---

1. The coefficient of $\bar{R}$ in the case of the maturity regression for bank 2 is misleading. A sampling difficulty caused the coefficient to be negatively biased. The conclusion in the text would probably still hold if this bias could be eliminated.

2. One banker suggested that this result reflected a diminution of demand for these funds. His argument is equivalent to saying that the partial loan offer function was not identified. Such arguments are not convincing unless reasons for lack of identification can be suggested.

When working capital, W_m and W_p, change in working capital $\left(\frac{dW}{dt}\right)_m$ and $\left(\frac{dW}{dt}\right)_p$, change in profits $\left(\frac{dP}{dt}\right)_m$ and $\left(\frac{dP}{dt}\right)_p$, and total assets H were individually added to the above regressions, the fits of the regressions were not significantly improved.[1] Therefore, hypotheses 6, 7, 8, and 9 are rejected.

Hypothesis 10 is accepted. The coefficient of the negative profit rate $\frac{P_m}{H}$ in the interest rate regression of bank 2 is significant. Firms which have negative profit rates are forced to pay higher interest rates. In the case of bank 1, the coefficient of $\frac{P_m}{H}$ in the security equation is significantly positive implying that more negative profit rates increase the likelihood of security. For the same equation of bank 2, the coefficient of $\frac{P_p}{H}$ is significantly negative implying that higher profit rates decrease the likelihood of security. Other coefficients of $\frac{P_m}{H}$ and $\frac{P_p}{H}$ are not significantly different from zero.

Hypotheses 11, 12, and 13 are rejected. Residuals of the above regressions were regressed on A_o, V_p, and E_a individually. Only those observations containing nonzero values of the independent variable (A_o, V_p, or E_a) were included in each of the calculations. Infrequent reporting of affiliated balances limited empirical examination of hypothesis 13 to one bank. None of the four slopes was significantly different from zero.[2]

Two banks reported the volume of participations A_o. In both cases no coefficients were significant. Interestingly, had the hypothesis implied opposite signs and the rejection regions been .05, both coefficients in the amount (residuals) equations would have been significant. In other words, simultaneous outside borrowing seems to increase the amount of term money a bank will lend to a borrower. One possible explanation is that confidence in a borrower grows in the mind of a banker if other bankers concur in his decision to lend. A second explanation, verbally supported by an officer in a bank, is that active participation in loan syndicates is a matter of prestige and a show of competitive strength. Further empirical work is necessary before these explanations can be accepted.

Variance of an applicant's profits was analyzed for each of the

---

1. The test of H was not precisely that described in the text. A covariance matrix containing Log P_p/H, Log P_p, and Log H is singular. When P_p/H and P_m/H were omitted from regressions, additions of H to the regressions did not significantly improve the fit. This was the basis for rejecting hypothesis 9.

2. It should be recalled that reporting of E_a was voluntary. More systematic data collection might conceivably alter the conclusion.

three banks and none of the coefficients was significant.

Two additional tests were performed on the term loan samples. The first examined the hypothesis that all three banks could be characterized by the same partial loan offer functions rather than separate functions for each bank. Employing the F test, the hypothesis is rejected at the .05 level in the cases of the rate of interest and maturity regressions. The ability of an applicant to obtain certain rates of interest and certain maturities clearly depends upon the bank to which he applies. The hypothesis is accepted in the cases of the amount and dummy security variables.

The second test examined the hypothesis that if the sample for each bank were split in two sections, loans granted from January 1955 through May 1956 and loans granted from June 1956 through October 1957, the estimated partial loan offer functions would not be significantly different from the equations reported in tables 2 - 2 through 2 - 5. Again employing the F test, the hypothesis is accepted in all four regressions of each of two banks. However, for a third bank, the hypothesis is rejected in the maturity and dummy security regressions at the .05 level.

In a comparison of the first and second half of the period, the following statements apply to that bank. Firms were required to have a higher current ratio in order to be granted a loan of a particular maturity. The bank became even less willing to grant long maturity loans as prime rose. A higher current ratio was necessary in order to achieve a particular probability of a security requirement. Further, the bank made fewer loans to firms which experienced mean losses over the preceding three to five years. A subsequent conversation with an officer of this bank confirmed that there was a sharp change in lending policy early in 1956. This change included a greater unwillingness to make term loans. It is of interest to observe that this tightening did not involve a change in the interest rate regression.

3. Analysis of Commercial and Industrial Lending

This section reports results of statistical tests performed on samples from the 1955 and 1957 surveys of business loans collected by the Board of Governors of the Federal Reserve System.[1] The surveys contain information on loans of various amounts and maturities, in contrast to the large, long maturity term loans considered in Section 2.

Data gathered in the surveys permit examination of the relevance of a few borrower characteristics to the loan offer function. More importantly, tests of hypotheses about the effects of some bank characteristics on bank lending can be made, because the data were collected from a large number of banks. Finally, tests of the impact of increased bond interest rates on bank lending can be performed. The last set of tests is

---

1. I am most grateful to the Board for making available summaries of these surveys. Valuable comments and assistance were obtained from Mr. Edward P. Snyder of the Banking Section of the Board's Division of Research and Statistics.

possible because samples collected in 1955 and 1957 were drawn at points of time when interest rates were respectively low and high.

3 - 1. Hypotheses

A description of the surveys and their sampling techniques may be found in the April 1956[1] and the April 1958[2] Federal Reserve Bulletins. For reasons of cost, only data relating to one district of the Federal Reserve System could be processed. Apparent balance in terms of industry, size of banks, and demand for loans suggested the Cleveland district and it was selected.

The surveys collected information on individual loans granted by a sample of member banks. Four conditions had to be met by the loans in the surveys if they were to be included in the present study:

(1) They must have been granted to firms in the United States (1955 only).

(2) They must have had maturities (when granted) of not more than ten years.

(3) They must have been granted or approved in the month immediately preceding the date of the survey.

(4) All variables considered in this section had to have been reported.

The following new variables were reported in the 1955 survey:

D = the level of deposits of the lending bank in June of the year of the survey. (in dollars)

L = ratio of lending bank's commercial and industrial loans to its total assets, as of the survey date. (in percent)

B = 10 if borrower was located in the same metropolitan area as the lending bank, B = 1 otherwise.

Both surveys reported H, the total assets of the borrower; R, effective loan rate of interest; M, the maturity of the loan when granted; A, the amount of the loan; and S, the dummy security variable. Unfortunately, the 1957 survey did not collect information on L and B.

Variables D, L, H, and M were recorded in class intervals. Thus the level of deposits of a lending bank might be reported in the interval of ten to twenty million dollars. If an observation fell into an interval, it was arbitrarily assigned the midpoint. If the class interval had an open end, an estimate of its mean was obtained from other sources.[3] This procedure may bias the results, but in most cases a sufficient number of classes

1. James B. Eckert, "Business Loans to Member Banks," The Federal Reserve Bulletin, XLII (April 1956), pp. 337-340.

2. James B. Eckert, "Member Bank Lending to Small Business, 1955-7," The Federal Reserve Bulletin, XLIV (April 1958), pp. 393-394 and pp. 410-411.

3. Details may be obtained by corresponding with the author.

existed to reduce this bias to negligible proportions. An exception was the loan ratio L where only four classes were available.

The theory developed in Section 1 assumed that loans were sufficiently large or of sufficiently long maturity that costs of lending could be ignored. In view of the fact that nearly forty percent of the loans in the Federal Reserve samples were to borrowers with total assets under $50,000 and ninety percent for periods of less than one year, this assumption may not be valid. Small loans and loans with short maturities are relatively less desirable than would be true if lending were costless. With this qualification,[1] the following hypotheses may be inferred from the theory:

(1) Ceteris paribus, as the total assets of a borrower increase, the rate of interest on the loan decreases and/or the maturity of the loan increases and/or the amount of the loan increases and/or the likelihood of security decreases.

(2) Ceteris paribus, as the level of deposits of a lending bank increases, the rate of interest on the loan increases and/or the maturity of the loan decreases and/or the likelihood of security increases.

(3) Ceteris paribus, as the loan ratio of the bank increases, the rate of interest on the loan increases and/or the maturity of the loan decreases and/or the amount of the loan decreases and/or the likelihood of security increases.

(4) Ceteris paribus, if the borrower resides in the same city as the lending bank, the rate of interest on the loan is lower and/or the maturity of the loan is longer and/or the amount of the loan is larger and/or the likelihood of security is smaller.

Supporting arguments for these hypotheses are developed in terms of R. The remaining relationships in the hypotheses may be inferred, mutatis mutandis, because R is believed to be a complement of M and A and a substitute for S.

Hypothesis 1 states that if two firms apply for a loan, then the firm with larger total assets will be charged a lower rate of interest. It may be defended on the grounds that bankers believe the probability of default is smaller when a borrower's assets are larger.

The fact that coefficients of borrower's assets were insignificant in Section 2 should not discourage use of the variable in this section. Profits and assets are highly correlated and the coefficients of profits were significant in Section 2. Information on a borrower's income was not obtained in the Federal Reserve surveys; total assets is viewed as a proxy for an applicant's profits.

Hypothesis 2 states that if identical firms apply to banks, differing only in size, then larger banks will charge higher rates of interest. This hypothesis is derived from three assumptions: (1) banks are averse to

1. This qualification is probably not serious. Small loans earn small sums, but their credit analysis is also simple and cheap.

deposit instability, (2) a bank's utility function is quadratic with respect to changes in the level of its deposits, and (3) the analysis in the appendix characterizes banks in the real world.

The hypothesis makes no statement about the relationship between the amount of the loan and the level of deposits of the lending bank. However, the relationship is probably positive because commercial banks are not permitted to lend sums in excess of ten percent of their net worth. Traditional bank examination standards require the level of a bank's deposits and its equity to be positively correlated.

Hypothesis 3 states that if identical commercial firms apply to banks, differing only in their business loan ratios, then banks with higher ratios of commercial and industrial loans to deposits will charge higher rates of interest.[1] This hypothesis utilizes the notion of portfolio balance. These banks have some common optimum portfolio; if the proportion of business loans is low, they lower rates to encourage more business lending. Rates will be raised if the banks have too many business loans.[2]

Specialization of lending may operate to offset this hypothesis. If some banks specialize in agricultural loans, they may feel quite inexperienced when lending to manufacturers or shopping centers. This lack of experience induces such banks to be conservative commercial and industrial lenders. Following the same line of reasoning, some banks may view themselves essentially as lending institutions while others prefer to be regarded as depositories. The former institutions are systematically more aggressive lenders. Their loans to borrowers of given means are at lower interest rates, of longer maturities, of larger amounts, and with lower probability of security. The latter banks are more conservative lenders.

The question of whether a common portfolio balance or specialization best describes the real world is important for appraising the ease of obtaining credit in certain communities where banks have high business loan ratios. There is no theoretical basis for selecting which of the two hypotheses is right. Empirical tests will doubtless reflect both patterns of behavior.

Hypothesis 4 states that if a bank receives two loan requests from borrowers identical in every respect, except that one of them is located

---

1. The reader may object that lending by banks during the month prior to the survey date may influence the loan ratio at the time of the survey. The ideal loan ratio for the purposes of this hypothesis is that existing a month prior to the survey date. However, this ideal statistic is not available and it is reasonable to assume that the ratio existing at the time of the survey had not been appreciably altered during the preceding month.

2. Cf. James Tobin, "Liquidity Preference as Behavior Towards Risk," The Review of Economic Studies, No. 67 (February 1958), pp. 65–86.

in a different city than that in which the bank and the other borrower reside, then the bank will charge a higher rate of interest to the first borrower. Among the reasons which support this hypothesis are: (1) higher costs of processing and checking distant borrowers, (2) a belief that the elasticity of demand for loans is greater in the bank's own city, and (3) greater ease of surveillance of the borrower.

3 - 2. Empirical Results

Partial loan offer functions were estimated from a sample of approximately 3000 loans. Hypotheses 1, 2, and 4 are one tailed tests of parameters and again the .025 level of significance was selected. Hypothesis 3 is a two tailed test and coefficients of L are tested at the .05 level.

As in Section 2, statistical tests of parameters in partial loan offer functions are not ideal tests of the theory. Hypotheses are again judged to be accepted if at least two coefficients are significant.[1] Table 3 - 1 reports the estimated partial loan offer functions.

All multiple regressions in table 3 - 1 are significant at the .01 level. Hypothesis 1 is accepted. Coefficients of total assets H are significantly different from zero in all four regressions. The asset elasticity of the rate of interest is .066. Thus if a firm had previously borrowed at eight percent and suddenly expanded its assets tenfold, it could expect to borrow at about seven percent.[2] Banks apparently do view a borrower's total assets as a measure of expected default risk. The reader should recall that total assets is a proxy for other income and balance sheet variables which were not recorded.

Hypothesis 2 is accepted; larger banks charge higher rates of interest to and are more likely to insist upon security from borrowers of a particular size. The failure of the hypothesis in the case of the maturity regression is a bit puzzling. One explanation is that small banks do not in practice make loans with long maturities. There is probably a tendency to issue short maturity notes with the implicit understanding that these notes will be renewed if conditions of the borrower do not change drastically. Larger banks, on the other hand, may prefer to make term loans because of the lower average legal and administrative costs associated with the larger loans which they grant. The failure of the coefficient to be significant may result from a failure of the survey to measure the effective maturity of loans of small banks.

Unfortunately, if L and B are omitted from the regressions, the coefficient of D is not significant in the interest rate equation. As is

1. The probability of a random variable being significant at the .025 level in at most one regression is .997.

2. The statement obviously makes the usual static cross-section assumption that differences between borrowers at a point in time are analytically identical to differences in a given borrower at different points in time.

TABLE 3 - 1

Partial Loan Offer Functions Estimated from the
1955 Federal Reserve Survey of Business Loans

Dependent Variable	R^2	Constant	Log D	Log L	Log H	Log B
Log R	.287	1.04771	.0054857 *	-.012094	-.066186 *	-.010048 *
			(.0023484)	(.0064626)	(.0020187)	(.0047189)
Log M	.009	.19964	-.021283	.11413 **	.022868 *	-.074805
			(.013009)	(.035799)	(.011182)	(.026140)
Log A	.513	-.056965	.060246 *	.088687 **	.63398 *	-.012915
			(.014878)	(.040943)	(.012789)	(.029896)
Log S	.018	.55192	.042968 *	.010879	-.081928 *	-.099366 *
			(.013516)	(.037194)	(.011618)	(.027158)

* implies significance at .025

** implies significance at .05

usually true when multicollinearity appears, the meaning of this second result is unclear. The hypothesis is still accepted on the basis of the significant coefficients, but further research is necessary before doubt is extinguished.

Hypothesis 3 is rejected in favor of the specialization hypothesis discussed above. The coefficient of the business loan ratio is significantly positive in the maturity and amount regressions and narrowly misses being significant in the interest rate regression. Banks with high business loan ratios make new business loans which have longer maturities and larger amounts to borrowers of any particular size. This result contradicts a view that banks have some common optimum portfolio toward which they constantly adjust.

An explanation for the failure of the hypothesis in the interest rate and dummy security variable regressions is that aggressive banks prefer to compete only with certain terms of loans, i.e., maturity and amount. This explanation is consistent with statements sometimes encountered in the banking literature concerning the undesirability of interest rate competition.[1]

Hypothesis 4 is tentatively accepted because of significant coefficients of B in the interest rate and dummy security variable regressions. The coefficient of B possesses the wrong sign in the case of the maturity and amount regressions. Indeed, if the hypothesis had predicted that the coefficient in the maturity regression be negative, the coefficient would have been significantly different from zero. No single explanation for the failure of the hypothesis in these regressions is obvious. Further research on the relationship of location to bank lending is needed.

3 - 3. The Impact of High Interest Rates

Higher interest rates on competing assets influence the terms at which firms may obtain loans. Analysis of the impact of different levels of interest rates on the willingness of commercial banks to lend can only be accomplished by examining bank lending at different points in time. If the effect of competing interest rate changes is to be identified, other variables which influence the willingness of a bank to lend must be considered at the various points in time. A brief discussion of factors influencing the loan offer function in October 1955 and October 1957 follows. It may be seen that, if anything, the net effect of these other factors would tend to discourage bank lending in 1957.

Both Federal Reserve surveys were conducted in early October, thus permitting seasonal determinants of lending to be ignored. Although a modest increase in unemployment was occurring in the fall of 1957, business expectations were by no means bleak. They were probably more pessimistic than in October 1955 when the economy was prospering. This change in the economic picture probably induced bankers to be more

---

1. Beckhart, op. cit., pp. 305–307.

cautious when granting loans. In both periods, discount rates had risen one half point in the two months preceding the survey dates. Very similar patterns were exhibited by the prime rate and bill rates in the same two month periods.

The most important difference in the loan market on the two survey dates was higher interest rates in 1957. Between October 1955 and October 1957 new issue bill rates jumped from 2.23 to 3.58, prime rose from 3.50 to 4.50, rates on bankers' acceptances rose from 2.23 to 3.75, and rates on United States Government long term bonds rose from 2.89 to 3.73.[1]

Another difference was that the ratio of loans to deposits of most commercial banks was higher in the second period. This change in bank portfolios would be expected to make banks more cautious when granting loans.

The following hypotheses may be inferred from the theory presented in Parts 3 and 4 of Section 1.

(5) Ceteris paribus, when rates of interest on competing assets are higher, borrowers of any particular size --
  (a) pay a higher rate of interest on their loans.
  (b) receive loans with shorter maturities.
  (c) receive smaller loans.
  (d) are more frequently required to provide security.

(6) Ceteris paribus, when rates of interest on competing assets are higher, the deposit coefficient (elasticity) of --
  (a) the loan rate of interest is more positive.
  (b) the maturity of the loan is more negative.
  (c) the amount of the loan is more negative.
  (d) the likelihood of security is more positive.

Part (a) of hypothesis 5 is consistent with the conventional theory that loan rates of interest clear loan markets. It may also be consistent with the "levels" version of the availability of credit doctrine which does not always prohibit movements in loan interest rates. But as observed earlier, supporters of that doctrine believe that these movements are insufficient to clear loan markets. Parts (b), (c), and (d) of the hypothesis reflect their view that credit rationing is necessary to clear loan markets when interest rates reach higher levels.

Credit rationing may also be achieved by increasing the number of loan rejections instead of altering terms of lending. No direct test of this form of credit rationing is possible in the present study because information on rejections is not available. However, it seems strange that bankers would not ration credit both by increasing rejections and modifying terms of lending. If only the former technique is employed, then banks must behave in the following fashion. If a firm seeks a two year loan of $10,000, the banker says "yes" or "no." He does not attempt to scale down the

1. Federal Reserve Bulletin, XLII (April 1956), p. 365, and Federal Reserve Bulletin, XLIV (April 1958), p. 457.

firm's request or to extract a faster repayment schedule. In fact, he behaves contrary to answers by bankers in a recent survey by the American Bankers Association. [1]

Hypothesis 6 implies that an increase in competing interest rates causes large banks to be even less willing to lend than small banks. Part (a) of this hypothesis reflects the conventional view that increased aversion to lending results in higher interest rates. Parts (b), (c), and (d) reflect the corresponding credit rationing arguments.

Table 3 - 2 reports multiple regressions of each of the terms of lending on the level of deposits of the lending bank, total assets of the borrower, and a shift constant. For purposes of testing the hypothesis, separate slopes were estimated for the variables in each of the two surveys. The total number of observations is on the order of 5000 and all regressions are significant at the .01 level.

Part (a) of hypothesis 5 is accepted. Inspection of the loan rate of interest regression reveals that in 1957 borrowers had to pay higher rates of interest. This conclusion is true because the coefficient of the shift constant is significantly positive and the asset elasticity of the loan rate of interest is significantly less negative. The change in asset elasticity implies in addition that the interest rate differentials between small and large borrowers were smaller in 1957. Acceptance of part (a) of hypothesis 5 is consistent with the conventional theory of the loan market and the availability of credit doctrine.

Parts (b), (c), and (d) of hypothesis 5 are rejected. First, coefficients of C_{57} are not significantly different from zero in the case of the maturity and dummy security variable regressions. In the amount regression, the coefficient of C_{57} is significantly positive implying that at least some borrowers could obtain larger loans in 1957, contrary to the availability argument.

Second, none of the three asset elasticities changed in such a way as to bear out the hypothesis. The 1957 asset elasticity of maturity is not significantly different from the corresponding 1955 elasticity. The asset elasticity of amount is significantly higher in 1957 than in 1955, contrary to the availability doctrine. A borrower of a given size could obtain larger loans in 1957 than in 1955. The asset elasticity of the dummy security variable is not significantly different in the two periods.

As observed previously, business conditions were deteriorating in October 1957. This deterioration should have changed the coefficients of C_{57} and H in precisely the direction suggested by the doctrine. The failure of the levels version of the availability doctrine is thus even

---

1. E. S. Adams, "Monetary Restraint and Bank Credit," Banking, L (September 1957), pp. 68–142.

TABLE 3 - 2

Partial Loan Offer Functions Estimated from the
1955 and 1957 Federal Reserve Surveys of Business Loans[1]

Dependent Variable	R^2	Constant	C_{57}	Log D_{55}	Log D_{57}	Log H_{55}	Log H_{57}
Log R	.255	.97374	.089367 *	.0033392	-.0013864	-.065526 *	-.037387 *
			(.0054164)	(.0021005)	(.0023366)	(.0019706)	(.0018213)
Log M	.007	.066015	.052545	-.0031113	-.0006963	.033151 *	.029188 *
			(.031816)	(.012339)	(.013726)	(.011565)	(.010698)
Log A	.591	-.24335	.22165	.074922 *	.070752 *	.63757 *	.68951 *
			(.034309)	(.013306)	(.014801)	(.012483)	(.011537)
Log S	.018	.50698	.0006333	.043566 *	.023140	-.071970 *	-.076586 *
			(.030988)	(.012018)	(.013368)	(.011274)	(.010420)

1. Definitions of the independent variables in Table 3 - 2 are as follows. A loan made in 1955 would have D_{55} = D, D_{57} = 1, H_{55} = H, H_{57} = 1, and C_{57} = 0. Similarly, a loan made in 1957 would have D_{55} = 1, D_{57} = D, H_{55} = 1, H_{57} = H, and C_{57} = 1.

* implies significance at the .025 level.

more conspicuous.[1, 2]

Between 1955 and 1957 bankers believed that credit rationing had not appreciably reduced the growth in their loans.[3] Their view is consistent with the outcome of tests of hypothesis 5.

Hypothesis 6 is also rejected. No deposit elasticity changed significantly in the two year period. In other words, there is no evidence that borrowers from large or small banks bear the burden of higher competing (bond) interest rates more heavily.

4. Estimation of the Loan Offer Function with Canonical Correlation.

This section reports attempts to estimate a commercial bank loan offer function with canonical correlation.[4] Substitution between terms of lending is explicitly considered when this technique is employed. One measure of the degree to which substitutability exists among terms of lending is the amount by which the canonical correlation exceeds the largest multiple correlation of the previous partial offer functions. Unfortunately, no probability statement can be made about whether the difference between these two correlations is significantly greater than zero.[5]

---

1. The regressions in table 3 - 2 suggest that the trade-offs between the loan interest rate and other terms of lending changed between the two surveys. Table 4 - 3 in the subsequent section reports that nonrate terms were relatively more important in 1957. But if all borrowers obtained credit at the same or better terms (with the exception of the loan rate of interest), this observation is of little value as an explanation of the effectiveness of monetary policy.
2. Professor Henry Wallich has verbally suggested an explanation of the above results consistent with the availability of credit doctrine. He argues that bankers normally tend to become more liberal lenders as the economy moves further away from a previous recession. If banks were unusually conservative lenders in 1955 because of the 1953-54 recession, then the arguments in the text must be qualified. I do not share this view.
3. E. S. Adams, loc. cit.
4. See p. 18.
5. The reader may object that this rather elaborate statistical procedure is just a ruse; examination of the partial correlation coefficients of the residuals of the partial offer functions would be equally satisfactory. The error in this objection is that partial correlation coefficients are not jointly determined with other parameters, as the theory requires. The signs of estimated partial correlation coefficients will depend critically on whether the function has been properly specified. Variables such as net worth and debt of a borrower have been omitted, thereby misspecifying the function. In fact, partial correlation coefficients were estimated for regressions reported in tables 2 - 2 through 2 - 5 and table 3 - 1. The signs of the coefficients were not in accordance with a priori expectations.

The theory in Section 1 suggests that for similar borrowers the loan rate of interest should increase as the amount or maturity of a loan increases and should be lower if security is supplied. Also, for similar borrowers the maturity of a loan should be shorter if the amount of the loan is larger and should be longer if security is offered. Finally, the amount of money a bank will be willing to lend to a particular borrower should increase if security is proffered.[1]

Table 4 - 1 reports canonical correlations for each of the three banks considered in Section 2.[2]

The hypotheses about terms of lending are supported by the reported coefficients (canonical weights) with the exception of the dummy security variable. That variable has limited economic meaning and it is not surprising that its canonical weight did not always possess the predicted sign.

The coefficients suggest that if a borrower with given characteristics requests a longer maturity loan, a more than proportionate reduction in amount will be necessary to compensate the bank, given other terms of lending. On the other hand, requests for either longer maturity loans or larger loans may be compensated for with a relatively small increase in the loan rate of interest. The rankings of the three coefficients are identical in the case of each of the three banks.

Among the borrower characteristics, only profit variables have the same sign for all three banks. Coefficients of bank 1 and bank 2 have identical signs with the exception of the profit rate variables. Table 2 - 5 reports that the latter are significantly related to S. The coefficients of S have different signs for the two banks.

In the cases of banks 1 and 2, coefficients of negative profits P_m and years as a depositor Y have signs which are not consistent with the previous hypotheses. Difficulties with P_m are again believed to be associated with misspecification. Coefficients of Y are not significant in the regression analysis for banks 1 and 2 and are interpreted to be spurious.

For bank 3, the coefficient of Y possesses the hypothesized sign, confirming the significance of this variable in the regression analysis. Because profit rates and balances are not significant in regressions for this bank, wrong signs are not surprising. The sign of the coefficient of the current ratio apparently derives from the fact that the coefficient of C is positive in table 2 - 2. Because rate equations tend to dominate the coefficients in table 4 - 1, the positive coefficient in table 2 - 2 outweighs the

---

1. The theory is informally tested in this section by the degree of conformity of signs of canonical weights to these a priori expectations.

2. Numbers in tables 4 - 1, 4 - 2, and 4 - 3 differ from those in true canonical equations by a constant of proportionality. As the analysis only concerns ratios of coefficients, comparisons will be more easily made if the equations are normalized. The value of the canonical correlation is invariant to such transformations.

TABLE 4 - 1

Canonical Correlations from Samples of Term Loans

Variable[1]	Bank 1	Bank 2	Bank 3
(Log R)'	1.0000	1.0000	1.0000
(Log M)'	-.053203	-.086959	-.070112
(Log A)'	-.012387	-.052804	-.032025
(Log S)'	-.015547	.000069	-.0013735
Canonical Correlation (Squared)	.799	.852	.796
(Log P_m)'	-.031719	-.022617	-.075585
(Log P_p)'	-.029489	-.052602	-.069635
(Log C)'	-.014137	-.011387	.018826
(Log E)'	-.0037316	-.0022580	.0049976
(Log Y)'	.0044043	.0027871	-.040074
(Log $\bar{R}$)'	.97548	1.12284	.91374
(Log P_m/H)'	-.0043322	.060639	.043162
(Log P_p/H)'	-.0049697	.0015566	.033827

1. Primes (') on variables indicate that they are defined as deviations from their means.

corresponding coefficient of C in table 2 - 3. No plausible explanation of the sign of this coefficient is obvious.

The squared canonical correlations exceed the largest previous R^2 by .031, .051, and .056 for banks 1, 2, and 3 respectively.

Table 4 - 2 reports canonical correlation results for the 1955 sample of loans collected by the Board of Governors of the Federal Reserve System.

The table demonstrates that all a priori predictions about trade-offs between terms of lending are verified. Again the coefficient of S should be regarded with some skepticism.

In the case of the Federal Reserve sample, increases in requested loan maturity by borrowers of a particular size are associated with less

TABLE 4 - 2

Canonical Correlations from Cleveland '55 Sample

Variable	Weight
(Log R)'	1.0000
(Log M)'	-.036209
(Log A)'	-.38171
(Log S)'	.066075
Squared Canonical Correlation	.547
(Log D)'	-.013906
(Log L)'	-.049349
(Log H)'	-.31442
(Log B)'	-.0089774

than proportionate decreases in loan amount, given other terms of lending. Apparently the reason for the difference between this result and the preceding term loan coefficients exists because the average business loan has a very short maturity. Doubling of average commercial and industrial loan maturities might involve extending maturities six months. In the case of term loans, however, a doubling of loan maturities might extend loans one to five years.

A substantial rise in the loan rate of interest is associated with requests for larger loans by borrowers of a particular size. This contrasts with the previous term loan result. Apparently interest rate differentials are more frequently used to compensate bankers for large risks in the commercial and industrial loan market than in the more restricted market for term loans. One reason for this behavior may be that term borrowers are relatively large firms having access to other sources of financing if bank rates rise too high. Consequently, banks obtain safety by reducing the maturity of loans to individual term borrowers. Other conditions of borrowing also compensate banks which do not charge large interest rate risk differentials. Term loan agreements usually require borrowers to keep some minimum level of working capital.

As in the case of term lending, only modest increases in interest rates are needed to compensate banks for increases in loan maturity, given borrower size and other terms of lending.

The coefficients of bank and borrower characteristics possess the hypothesized signs with the exception of deposits of the lending bank, D. As explained in the previous section, D is necessarily correlated with A because of maximum legal lending limits established by bank inspectors. Because the amount regression tends to dominate the canonical equation, the coefficient of D has the negative sign.

The increase in the squared canonical correlation over the R^2 of the amount regression was .034.

TABLE 4 - 3

A Comparison of Canonical Correlations Computed from Cleveland 1955 and Cleveland 1957 Samples

Variable	Weights 1955	Weights 1957
(Log R)'	1.00000	1.00000
(Log M)'	-.033975	.011718
(Log A)'	-.38626	-1.67200
(Log S)'	.065991	.46584
Squared Canonical Correlations	.546	.664
(Log D)'	-.041459	-.170110
(Log H)'	-.58216	-1.90516

Table 4 - 3 presents comparable canonical equations estimated from Cleveland district member banks in the 1955 and 1957 Federal Reserve surveys. There appear to have been three interesting changes between the two surveys. First, the weight placed on the loan interest rate relative to other terms of lending declined, consistent with the change in the asset elasticity reported in table 3 - 2. I suspect that some advocates of the availability of credit doctrine perceived this relative change and inferred that this was the mechanism by which monetary policy became effective. However, as reported in Part 3 - 3, this inference is incorrect because there is no evidence that any size of borrower was forced to accept either a smaller loan, a

shorter loan,[1] or a higher frequency of security in 1957 than in 1955. Possible explanations of this change include proximity of rates to maximums specified by usury laws, growing criticism of high loan rates, etc.

Second, the canonical correlation was substantially higher in 1957 than in 1955. Lending behavior of banks could be more precisely predicted in 1957 from a knowledge of an applicant's total assets and a bank's level of demand deposits. One possible explanation of this greater precision is that, as interest rates rose and loan portfolios grew, banks made fewer loans for noncredit reasons, i.e. loans motivated by competition for a new customer or loans to old "friends of the bank," etc.

Finally, the weight on loan maturity changed signs. As the weight on this variable in both equations is very small, this result probably should not be regarded too seriously. If the change actually occurred, it implies that banks viewed maturity as a desirable property. One explanation is that, if banks expected loan interest rates to drop in the near future, they could improve their future earnings by lending at longer maturities. Further evidence is necessary before this conjecture can be accepted.

The squared canonical correlations exceeded the corresponding R^2 of the amount regressions by .034 and .018 in 1955 and 1957 respectively.

5. Summary and Policy Implications of the Analysis

The research is summarized in Part 1 of this section. Policy implications of the analysis are reported in Part 2, and Part 3 suggests the relevance of the findings to the availability of credit doctrine.

5 - 1. Summary of the Analysis

In Section 1 a theory was proposed to explain the terms at which a bank with particular characteristics would lend to a loan applicant.

Certain simplifying assumptions were made in order to develop the theory. First, it was assumed that broad asset choices between bonds, loans, and cash could be generally ignored. Second, it was assumed that the services of banks as consultants or sources of information could be ignored. Third, deposit structure, equity, management, etc. of any bank were assumed to be fixed. Finally, the loan market was assumed to be imperfect.

A loan offer function was defined to be a relation which specifies the terms at which a bank with particular characteristics will be willing to lend to a borrower with a known profit, balance sheet, and credit history and particular prospects for the future.

It was then possible to speak of characteristics of an applicant and characteristics of a bank defining some maximum or efficient set of terms of lending. Four principal terms of lending were considered: (1) the loan rate of interest, (2) the loan maturity, (3) the amount of the loan, and

1. The coefficient of $\bar{R}$ in maturity regressions in Section 2 is an exception. This point is considered again in Part 5 - 3.

(4) the likelihood of security. Effective bargaining by borrowers was assumed to force bankers to this efficient frontier. Banks were hypothesized to be willing to substitute among terms of lending. Thus longer maturity loans might be obtained by paying higher interest rates and/or accepting smaller amounts of money and/or by offering security. Relationships between various borrower and bank characteristics and terms of lending were hypothesized.

The various hypotheses were tested with multiple regression and canonical correlation analysis. Five samples of data were analyzed. Three of the samples were drawn from term loans granted by three large commercial banks during the period from January 1955 through October 1957. Two other samples were extracted from the massive commercial and industrial loan surveys conducted by the Board of Governors of the Federal Reserve System in October 1955 and again in October 1957.

An applicant's profits, the ratio of his current assets to current liabilities, his deposit balances, the number of years which he was a depositor at the bank, his profit rate, his location, and perhaps his total assets appear to influence significantly the terms at which banks lend. The size of banks (a proxy for deposit variance), the bank's ratio of commercial and industrial loans to deposits, and the level of the prime rate of interest also affect the terms at which banks lend.

Section 4 reported that with canonical correlation analysis, it appears that banks do trade off terms of lending in the manner specified by the theory. When a borrower with particular characteristics seeks a longer maturity loan, he must expect either to pay a higher rate of interest and/or receive a smaller amount. The trade-off between these terms of lending and the dummy security variable was not consistently in the direction specified by the theory. In part this failure is associated with the crudeness of the security measure.

5 - 2. Implications of the Analysis for Monetary Policy

Policy implications of the analysis are considered from two points of view: (1) economic efficiency at a point in time and (2) the impact of monetary policy.

It is assumed economic efficiency implies that banks employ their funds so as to obtain a high rate of return and a low probability of bank failure. It is trite to observe that uncertainty of future deposit levels and uncertainty of loan repayment complicates this maximization problem.

There is no a priori reason for doubting that statistics such as a borrower's level of profits, his current ratio, and his profit rate successfully discriminate between safe and risky loans. Further, if the future rate of return of investments can be predicted by past profit and profit rate variables, banks which use these criteria cause capital to flow into high-yielding investments.

Recent proposals have been made to permit commercial banks to pay interest on demand deposits.[1] It has been argued in Section 2 that

balances have the dual properties of increasing effective interest rates and reducing deposit variance. The amount regressions in Section 2 support this argument. The size of term loans granted by banks 1 and 2 to a borrower are an increasing function of his past demand deposit balances. If banks prefer to hold smaller deposit balances as a result of being forced to pay interest on demand deposits, they will not care to make as many large term loans. The proposal must be considered from this perspective also.

It is not obvious that the number of years which a borrower has carried deposits with a bank is a socially desirable criterion for lending. Why is this variable considered? Information about borrowers is not costless. If borrowers supply more information to banks by maintaining a relationship, they have reduced a bank's cost. Criticism of this inertia by bankers is more a criticism of imperfect information flows than a comment on bank efficiency. Eliminating personal relationships between banks and borrowers is not desirable unless some other method of conveying information can be demonstrated to be cheaper.

Although no appraisal of the effectiveness of monetary policy on the level of economic activity can be attempted, the analysis does permit some discussion of the impact of monetary policy on bank lending.

Results in Section 2 suggest that increases in the prime rate of interest are associated with a shortening maturity of new term loans. Because prime tends to move with other interest rates in the economy, this implies that when monetary authorities permit rates to rise, they are in fact sanctioning a reduction in long term loans. If capital markets are imperfect, these long term borrowers may delay or may not be willing or able to carry out their projects. There arises a definite question as to whether such changes in the capital market are consistent with the objectives of monetary policy.

Analysis of the commercial and industrial loan samples in Section 3 indicated that although interest rates on these loans rose with other interest rates over the period 1955–57, no other terms of lending were made more restrictive. It was observed that the asset elasticity of the loan rate of interest became significantly less negative over the period. Again a change in capital markets resulted when other interest rates were permitted to rise.

The point, then, is that monetary policy does alter the composition of flows through the capital market. Until recently, very little attention has been directed to the character of these changes. The distinction between monetary policy and discretionary governmental controls is not very sharp. Decisions to employ monetary weapons rather than discretionary controls must be evaluated in terms of both the changes in the allocation of resources and the resulting reduction or expansion in aggregate demand.

---

1. Cf. James Tobin, "Towards Improving the Efficiency of the Monetary Mechanism." The Review of Economics and Statistics, XLII (August 1960), pp. 276–278.

5 - 3. Implications of the Analysis for the Availability of Credit Doctrine

The availability literature[1] suggests that either rising or high levels of interest rates result in credit rationing. No test of short term changes in interest rates could be made in this study. Because the level of interest rates was substantially higher in 1957 than in 1955, the comments of this paper pertain to the "levels" rather than "change" versions of the availability doctrine.

There is some evidence that banks are less willing to grant long maturity term loans when interest rates are higher. Because long maturity loans are inherently more risky, this result might be interpreted as credit rationing. Whether this behavior derives from changes in competing interest rates, higher levels of competing interest rates, or simply a growing proportion of loans in a bank's portfolio cannot be discerned.

Analysis of commercial and industrial loans revealed that borrowers of any particular size could obtain loans of at least the same maturity, the same amount, and with the same likelihood of security in 1957 as in 1955. If risk is associated with longer maturity, greater amount, or no security, we have the result that identically risky loans could be obtained by borrowers of given means, as measured by their total assets, regardless of the level of competing interest rates. There was no credit rationing![2]

This behavior is not consistent with the "levels" version of the availability of credit doctrine unless bankers behave in the following peculiar fashion. When a man comes to a bank with a loan request, the banker says "yes" or "no"; there is no substitution among terms of lending. Bankers do not scale down loans or obtain shorter maturities.

In order to test the availability doctrine ultimately, detailed information on loan rejections must also be studied. Unfortunately, this is difficult because a loan rejection is not well defined.

It has been observed earlier,[3] with respect to an American Bankers Association survey, that most bankers felt increased selectivity of lending had not appreciably restricted growth of their lending. This conclusion is supported by a study of large samples of commercial and industrial loans reported in Section 3. Although some evidence in support of availability arguments was exhibited in the rather specialized case of term lending, it is doubtful that great reliance can or should be placed on credit rationing.

---

1. See Section 1, Part 4.

2. As reported in Section 4, there is evidence that bankers were less willing to substitute between the loan rate and other terms of lending. But this change in the relative roles of terms of lending is not an explanation of the effectiveness of monetary policy.

3. See p. 44.

Appendix

An important factor in predicting bank lending is the instability of bank demand deposits. An empirical conclusion of this appendix is that the ratio of the variance of demand deposits to the level of deposits is an increasing function of bank size. Further, growing banks or declining banks will have a smaller ratio than banks which are of unchanging size. Three assumptions are necessary to demonstrate these conclusions.

First, assume that the rate at which transactions (deposits and withdrawals) occur can be represented as a non-linear function of the level of a bank's demand deposits.

$$(1) \qquad \frac{dn}{dt} = \alpha D^{\gamma} \qquad \alpha, \gamma > 0 .^{1}$$

where:

$\frac{dn}{dt}$ = number of transactions per unit time.

D = a bank's demand deposits net of interbank and government deposits.

Second, assume that the size of the average transaction is independent of whether it is a deposit or withdrawal and that it is also a non-linear function of the level of demand deposits.

$$(2) \qquad s = \beta D^{\delta} \qquad \beta, \delta > 0 .$$

where:

s = the size of the average transaction.

Finally, assume that transactions are distributed according to the binomial distribution with p = the probability that the next transaction will be a deposit, 1 - p = the probability that it will be a withdrawal. Growing banks have $p > .5$ and declining banks have $p < .5$.

The expected value of the level of deposits at the close of a period T units long is:

$$(3) \qquad D_{t+T} = D_t + (2p - 1) \alpha \beta D^{\delta+\gamma} \cdot T .$$

---

1. Very recently, some evidence has been published suggesting that $\gamma = 1$. If so, the conclusion of this appendix is considerably strengthened. $\gamma = 1$ permits the conclusion that the ratio of the standard deviation of deposit level to the level of deposits is an increasing function of bank size. Then the hypotheses in the text may be restated in terms of the standard deviation of deposit level instead of variance of deposit level. Cf. "Interpretation of Size-Cost Relationships in Banking," Monthly Review of the Federal Reserve Bank of Kansas City, March 1961, pp. 8-9.

The variance of the distribution of changes in deposits is given by:

$$\sigma^2 = 4T \cdot (p - p^2) \quad \alpha\beta^2 \; D^{\gamma+2\delta} \; . \tag{4}$$

A number relevant to this discussion is the turnover rate of demand deposits. In terms of the above, we must express the expected turnover as:

$$R = (1 - p) \quad \alpha\beta D^{\gamma+\delta-1} \cdot T \; . \tag{5}$$

$$\frac{\partial R}{\partial D} = (1 - p)\, \alpha\beta \quad (\gamma + \delta - 1)\, D^{(\gamma+\delta-2)} \cdot T \; . \tag{6}$$

Because T, (1 - p), α, and β are positive, a necessary and sufficient condition for this derivative to be positive is:

$$\gamma + \delta > 1 \; . \tag{7}$$

Multiplying (4) by 1/D and differentiating with respect to deposits yields:

$$\frac{\partial \frac{\sigma^2}{D}}{\partial D} = 4(p - p^2)\, \alpha\, \beta^2 \, (\gamma + 2\delta - 1)\, D^{\gamma+2\delta-2} \cdot T \; . \tag{8}$$

Because $(p - p^2)$, α, and T are positive, a necessary and sufficient condition for (8) to be positive is that $\gamma + 2\delta > 1$. Because $\delta > 0$, if (7) holds the result is proven.

TABLE 1

Average Demand Deposit Turnover Rate, 1956[1]

Banks in	R
(a) New York City	45.8
(b) 6 other centers[2]	28.8
(c) 337 other reporting centers	21.8

1. The Federal Reserve Bulletin, XLIV (April 1958), p. 446.
2. Boston, Philadelphia, Chicago, Detroit, Los Angeles, San Francisco.

It is well known that the average bank size in New York City, in terms of volume of deposits, is the largest in the United States. It is also quite clear that the average bank size in the six other large centers exceeds that in

337 other reporting centers. Assuming all other things are equal,[1] this implies that (7) holds.

Consequently, the first result is demonstrated. The second result follows directly from (8) because $(p - p^2)$ will be a maximum when $p = 5$.

Introducing stochastic terms in equations (1) and (2) does not seem to qualify seriously the conclusion of this appendix. The result will depend upon the sign and magnitude of the covariance of these terms. There does not appear to be an obvious a priori reason for rejecting the assumption that their covariance is zero.

Finally, no cognizance of the important retention rate of transactions by a bank has been made in this appendix. If the bank being considered is a monopoly bank, then all the withdrawals will later appear as deposits and the instability of deposits reduces to fluctuations in the bank's float. On the other hand, if the bank expects that all checks drawn on it will never again appear as deposits, then variance of deposits is critical. It is assumed that size of bank is not correlated with the retention rate of transactions.

---

1. In particular, it is assumed that there are no aggregation effects in the consolidated turnover statistics. This means that a bank of a particular size in New York will have the same turnover experience as a similar sized bank in another city. There is no obvious way to test this assumption without additional evidence.

Summary: This essay is a theoretical investigation of the problem of deriving the criteria which should govern the choice among alternative techniques of production of varying capital intensities. A simple two-sector dynamic model of an economy is considered. Sector I produces machines and Sector II produces consumer goods using machines and labor. Alternative techniques of production are available within Sector II, each characterized by the labor intensity associated with the type of machines it uses. The labor force is assumed to grow at a constant geometric rate. It is shown that the economy considered is capable of generating a maximum sustainable rate of consumption per worker. This rate is attained at each point of time along a "terminal path" in which only one of the many alternative techniques of producing consumer goods is used. Approaches to the "terminal path" from an initial position not on it are also considered. In particular, the time path which minimizes the time needed to reach the "terminal path" from below is derived. The growth paths that result from following some of the investment criteria proposed in the literature on economic development are examined. Some characteristics of the class of all efficient growth paths are also derived.

This essay was submitted to Yale University in 1961 in partial fulfillment of the requirements for the degree of Doctor of Philosophy, to be awarded in 1962. The author is currently Assistant Professor of Economics at Yale University and on the staff of the Cowles Foundation for Research in Economics at Yale.

Dissertation Committee: Professors Tjalling C. Koopmans (Chairman), William J. Fellner, and John W. Hooper.

*

I wish to thank Professors Tjalling Koopmans, William Fellner, Alan Manne and John Hooper for their valuable comments on earlier drafts of this essay. Thanks are also due to Mrs. Karen Hester and Miss Nancy Bowles for stylistic improvements. The Ford Foundation financed part of this research. The remainder was conducted as part of the Cowles Foundation research program, supported by the National Science Foundation and by the Office of Naval Research. None of the people mentioned above are in any way responsible for any errors remaining in this essay.

T. N. S.

INVESTMENT CRITERIA AND CHOICE OF TECHNIQUES OF PRODUCTION

Thirukodikaval N. Srinivasan

1. Introduction

Recent literature on economic development has devoted considerable attention to the problem of deriving the criteria which should govern the choice among alternative techniques of production. Widely differing solutions have been proposed leading to widely differing choices, from the most capital intensive to the most labor intensive techniques of production. These differences result in part from differing assumptions about the economic structure of underdeveloped areas and in part from errors in logic.

The problem of investment criteria is a sub-problem of the more general problem of choice among alternative time paths attainable by an economy from some given initial and anticipated future conditions. The choice is made by maximizing some social welfare function. In evaluating alternative time paths, some authors have restricted themselves to a finite time horizon and ignored the effects of their choice beyond this horizon. One has to keep in mind the different social welfare functions and the different time horizons of the various authors in comparing their proposals.

The literature on this topic is vast and an excellent summary can be found in a paper by Hollis Chenery.[1] For our purposes it is sufficient to take up what we consider as the most important of the various proposals: (a) the Social Marginal Productivity (SMP) criterion of A. E. Kahn and Hollis Chenery, (b) the Marginal Per Capita Reinvestment Quotient (MRQ) criterion of Walter Galenson and Harvey Leibenstein, (c) the Marginal Growth Contribution (MGC) criterion of Otto Eckstein, and (d) the Reinvestible Surplus criterion of Maurice Dobb and A. K. Sen.[2]

The social welfare function implicit in Kahn's work and made explicit by Chenery is a function of several variables, of which Chenery

---

1. Hollis B. Chenery, "Comparative Advantage and Development Policy," American Economic Review, LI (March 1961), pp. 18-51.

2. The relevant articles and books are: A. E. Kahn, "Investment Criteria in Development Programs," Quarterly Journal of Economics, LXV (February 1951), pp. 38-61; Hollis B. Chenery, "Application of Investment Criteria," Quarterly Journal of Economics, LXVII (February 1953), pp. 76-96; Walter Galenson and Harvey Leibenstein, "Investment Criteria, Productivity, and Economic Development," Quarterly Journal of Economics, LXIX (August 1955), pp. 343-370; Otto Eckstein, "Investment Criteria for Economic Development and the Theory of Intertemporal Welfare," Quarterly Journal of Economics, LXXI (February 1957), pp. 56-85; Maurice Dobb, Economic Growth and Planning (1st ed.; New York: Monthly Review Press, 1960), Chs. 3 and 4, pp. 29-75; A. K. Sen, Choice of Techniques, An Aspect of the Theory of Planned Economic Development (1st ed.; Oxford, England: Basil Blackwell, 1960), Chs. 2 and 3, pp. 21-51.

considers national income and balance of payments to be the most important. The problem they solve is the allocation of a given rate of investment of one period between alternative projects in such a way as to maximize the average annual increment in social welfare. This annual average is to be computed through a discounting process; the way in which the discount factor is to be chosen is not explained. Chenery's procedure is to rank projects according to their SMP and choose projects in decreasing order of SMP until the given rate of investment is exhausted. The SMP of a project is defined by Chenery as the average annual increment in national income (plus balance of payments effect) resulting from the marginal unit of investment in that project. The increment in national income is to be computed after applying some corrections to market prices which take into account the divergence between social and private benefits that occur when external economies, unused resources, tariffs, and subsidies are present.

In summary, the Kahn-Chenery approach is static since the choice variables belong to a single period. In computing the increment to the flow of national income resulting from the given investment, planners need to know the future product and factor prices. But future relative scarcities of various factors and products will themselves depend on the current rate and composition of investment. The same remark applies to the corrections with which Chenery wants to alter the market prices. By taking a static view Kahn and Chenery are implying either (a) that the projected investment is too "small" to affect the future course of prices, or (b) that the future course of prices is independent of current choices. Either assumption is untenable in the context of economic development.

Galenson and Leibenstein also criticize Chenery for using tools of economic statics in analyzing the problem of economic growth. The social welfare index they propose instead is an increasing function of the per capita output potential at some future point in time. The per capita output potential at any point in time depends, in their view, on a number of factors, of which capital per worker and the quality of the labor force are important. Capital per worker at any future point in time T depends on the accumulation of capital from now until T and the size of the labor force at T. Galenson and Leibenstein conclude that the "best allocation of investment resources is achieved by equating the MRQ of capital in its various alternative uses."[1] The MRQ of a project is determined by the annual surplus it generates over and above wage and depreciation costs, taking into account the contribution of the project toward improving the quality of the labor force and bringing about a decline in its rate of growth. Galenson and Leibenstein come to an empirical judgment that the MRQ is likely to be higher in capital intensive than in labor intensive projects.

Galenson and Leibenstein have been criticized by John Moes,[2]

1. Galenson and Leibenstein, op. cit., p. 351.

2. John Moes, "Investment Criteria, Productivity and Economic Development: Comment," Quarterly Journal of Economics, LXXI (February 1957), pp. 161-164.

Francis M. Bator[1] and others for the extreme nature of the assumptions needed to validate their theoretical and empirical conclusions. Even if one grants their assumption that only profit earners save, it is by no means necessary that profit maximization should always imply maximization of the capital labor ratio. Their prescription that governments in underdeveloped countries should maintain artificially high wages to induce producers to choose high capital labor ratios is questionable. If one is interested in retarding the rate of growth of the labor force, investing in capital intensive projects in a labor surplus economy is surely an indirect and perhaps an inefficient method. Lastly, the objective of maximizing the per capita output potential at some future point in time implies that a one commodity economy is being considered. This criterion is not very meaningful in an economy with more than one commodity. We shall come back to this point in Section 4.

Eckstein tries to synthesize the Kahn-Chenery approach and the Galenson-Leibenstein approach. His social welfare function is the sum of the discounted value of the stream of consumption resulting from a given current investment $\overline{K}$ and the future reinvestments occasioned by it. Current investment is to be divided among alternative projects which differ in their output stream and reinvestment potential. All reinvestments (at future points in time) are to be directed to a single project. Maximizing his social welfare function, Eckstein arrives at his MGC criterion. The optimal allocation of $\overline{K}$ results when the amounts invested in each project are such that the MGC's of the different projects are equal. The MGC of a project is the sum of two terms, (a) the present value of the project's direct contribution to the consumption stream and (b) the present value of the consumption stream resulting from reinvestments associated with the project. Eckstein also tries to determine the rate of investment itself by determining the interest rate through a marginal utility approach.

Eckstein's approach is subject to a number of criticisms. First, if one wants to consider an infinite time horizon as Eckstein does, one cannot proceed as if capital were the only scarce factor at all points in time. Second, there is no reason to direct all reinvestments to one project. One can, at the very least, postulate that the same projects are available at all future periods of time as at the beginning.[2]

The approaches of Dobb and Sen are very similar and we shall discuss only that outlined by Sen. Sen considers an economy consisting of two sectors, an advanced sector and a backward sector. The backward sector produces a single consumer good, corn, presumably using only

---

1. Francis M. Bator, "On Capital Productivity, Input Allocation and Growth," Quarterly Journal of Economics, LXXI (February 1957), pp. 86–106.

2. Eckstein is aware of this problem but he does not offer a complete solution. (Eckstein, op. cit., p. 72.)

labor. The advanced sector about to be initiated consists of two departments, one which produces corn, using labor and machines, and one which produces machines, using only labor.[1] The corn productivity of labor in department 1 depends on the capital intensity of the process used. The capital intensity of a process is measured by the number of man years required in department 2 to produce enough capital (machines) to equip one laborer in department 1. The wage rate to be paid to laborers in the advanced sector is assumed given. The total wage bill in the advanced sector is to be met from corn production within the sector except in the first period, when the advanced sector is being initiated. The wage bill of the first period is to be met by an amount of corn surplus extracted from the backward sector.

Sen considers three possible objectives for his economy: (1) maximizing the corn output of the second period in the advanced sector, (2) maximizing the rate of growth of corn output, and (3) maximizing the undiscounted sum of the stream of output of corn over T periods of time. Naturally the optimal capital intensities for attaining the three objectives are different.

One can agree with Sen in dismissing the first objective as clearly not the one to be pursued if one is interested in the problem of growth, for maximization of total corn output need not necessarily result in maximum surplus (over wages) for reinvestment unless the wage rate is zero.[2] One can even construct examples in which maximization of corn output of the second period yields a negative surplus. But the second objective, maximizing the rate of growth of corn output by maximizing the rate of surplus in each period, is based on the assumption that any rate of growth can be sustained indefinitely by withdrawing labor from the backward sector without reducing its output. Once we introduce a limit to the withdrawal of labor from the backward sector, the optimal capital intensity under the second objective may be different. The third objective completely ignores the time path of consumption by considering only the aggregate consumption over an interval of time. Furthermore, such an

---

1. Sen relaxes the assumption that capital goods can be produced only with labor in Chapter III of his book. But he does not go beyond stating that a solution to the problem of choice of capital intensity in the two departments is not impossible. (Sen, op. cit., Ch. III.)

2. Let $\bar{s}$ be the surplus corn extracted from the backward sector in the first period. Assume with Sen a lag of one year in the production of machines and no lag in the production of corn. Let a be the capital intensity chosen and w the wage rate. Then the total output of corn in the advanced sector in the second period is $\frac{\bar{s}}{wa} f(a)$ where $f(a)$ is the corn productivity of labor when capital intensity a is used. But surplus over wages is $\frac{\bar{s}}{wa} [f(a)-w]$. The maximization of one is not equivalent to that of the other.

approach precludes consideration of an infinite time horizon because undiscounted aggregate consumption over an infinite time horizon may well be infinite.

Thus our brief review of the more important articles in the literature reveals that most of these authors analyze the problem of growth in a static framework. Even the analysis of Galenson and Leibenstein, considered by them to be dynamic, is static in our view because they limit themselves to the objective of maximizing per capita output potential at some chosen point of time in the future. They therefore have a finite time horizon and hence static tools could be used to analyze their problems. Sen and Dobb do introduce dynamic elements explicitly in their model, but they fail to consider restraints, such as the availability of labor, that may become operative at future points in time. We shall attempt in the next section to develop a dynamic model which explicitly introduces constraints operative at each point in time and contains some essential elements of choice inherent in growth problems. We shall consider only two constraints (other than the nonnegativity restraints) at each point in time. The first one limits the employment at any point in time to the available labor force. The second limits the rate of gross investment in any period to the rate of output of machines in that period. This choice of constraints does not imply that these are the only important constraints relevant to a problem. In fact, constraints such as the availability of foreign exchange, skilled labor, and entrepreneurial skills may be equally important. Our choice of constraints is motivated purely by considerations of analytical convenience and our desire to analyze the problem of growth in a simple setting which still retains some essential elements.

In Section 2, we shall also be concerned with the derivation of the growth path called the "terminal path," which yields the maximum sustainable rate of per capita consumption (this concept will be defined later) at each point in time, assuming that the initial conditions are subject to choice.

In Section 3, we shall relax the assumption that the initial conditions are subject to choice and study the approach to the terminal path from historically given initial conditions. In particular, the growth path that minimizes the time taken to reach the terminal path will be derived.

In Section 4, the consequences of following an objective of maximizing the per capita consumption attainable at some point of time in the future, starting from given initial conditions, will be examined.

In Section 5, the objective of maximizing the discounted sum of the future stream of per capita consumption obtainable from given initial conditions will be studied. Section 6 will be concerned with characterizing the class of all efficient time paths (to be defined later). Section 7 will summarize and qualify some of the conclusions of the previous chapters.

Each of the following five sections will be divided into two subsections. The first sub-section will introduce the problem to be considered and present the solution. The second sub-section will be devoted

to detailed proofs. The reader interested mainly in the results can omit the proofs without losing continuity.

2. The Model and the Terminal Path

2.1. Introduction and Results

We shall discuss the choice of techniques of production in terms of the following model. Consider an economy consisting of two sectors or departments: Sector I produces machines of various types and Sector II produces a homogeneous consumer good. Within Sector II there are alternative techniques of production, each technique characterized by the particular combination of labor and machines it uses. A machine associated with one technique of production in Sector II cannot be used with any other technique of production. In Sector I, one type of machine is employed with a fixed labor input to produce any machine. Thus any machine used in either Sector I or Sector II may be produced from this combination of inputs. This assumption enables us to use a common unit of measurement for various types of machines prior to their construction while retaining the assumption that once constructed, machines of various types are not substitutes for each other. Constant returns to scale are assumed to prevail in both sectors. All types of machines once constructed are assumed to depreciate physically at the same geometric rate per unit of time. The labor force is assumed to grow at a constant geometric rate per unit of time.

The following are the variables, notations, and abbreviations used throughout this essay:

Class of Variable	Variable	Notation[1]	Dimension[2]
Time[3]	time	t	t
Given constants	geometric rate of growth of the labor force	Θ	$1/t$
Given functions	available rate of labor services at t	$L(t)$	ℓ/t

1. Subscript 0 is associated with the technique of producing machines. Subscript i (i = 1, 2, 3, ...) is associated with the i^{th} technique of producing consumer goods.
2. In this column, t = time, c = unit of the consumer good, m = unit of machines, and ℓ = unit of labor services.
3. Both continuous and discrete time versions of the model are considered. In the continuous version rates refer to instantaneous rates and in the discrete version to rates per period. In the discrete version it is convenient to assume that the inputs in any period make available outputs that can be used in the next period.

Class of Variable	Variable	Notation	Dimension
Technological constants	rate of input of labor services needed to operate a machine of type i i = 0, 1, 2, ...	λ_i	ℓ/m/t
	rate of output of machines (of any type) per machine of type 0 in use	γ_0	1/t
	rate of output of consumer good per machine of type i in use i = 1, 2, ...	γ_i	c/m/t
	geometric rate of depreciation of any type of machine	δ	1/t
Policy variables	stock of machines of type i in existence at time t i = 0, 1, 2, ...	$K_i(t)$	m
	stock of machines of type i in use at time t i 0, 1, 2, ...	$K_i'(t)$	m
	rate of gross addition to the stock of machines of type i at time t i = 0, 1, 2, ...	$X_i(t)$	m/t
	rate of output of consumer goods at time t	C_t	c/t
	rate of consumption per worker at time t	c_t	c/ℓ/t
Abbreviations	maximum sustainable rate of consumption per worker	MSCP	c/ℓ/t
	sustainable rate of consumption per worker	SCP	c/ℓ/t

It is clear that by assumption $L(t) = L(0)[1 + \theta]^t$ (or $L(0)e^{\theta t}$ if t is treated as a continuous variable). Without loss of generality we can assume that $L(0) = 1$ and that the parameters of the technology of the consumer goods industry satisfy the following:[1]

$$\frac{\lambda_1}{\gamma_1} > \frac{\lambda_2}{\gamma_2} > \frac{\lambda_3}{\gamma_3} > \ldots > \frac{\lambda_i}{\gamma_i} > \frac{\lambda_{i+1}}{\gamma_{i+1}} > \ldots\ldots \tag{2.1}$$

$$\gamma_1 > \gamma_2 > \gamma_3 > \ldots > \gamma_i > \gamma_{i+1} > \ldots\ldots \tag{2.2}$$

$$\frac{\gamma_1 - \gamma_2}{\lambda_1\gamma_2 - \lambda_2\gamma_1} < \frac{\gamma_2 - \gamma_3}{\lambda_2\gamma_3 - \lambda_3\gamma_2} < \frac{\gamma_i - \gamma_{i+1}}{\lambda_i\gamma_{i+1} - \lambda_{i+1}\gamma_i} < \ldots\ldots \tag{2.3}$$

(2.1) and (2.2) state that the techniques have been numbered according to decreasing labor intensity.[2] (2.3) ensures that those techniques which require more labor and machines per unit of output than a suitable combination of other techniques have been eliminated from the set of alternative techniques. Figure 2a shows an isoquant of consumer goods production. This isoquant shows the alternative combinations of machines and labor that can be used to produce a rate of output of one unit of consumer good per period. It is worth emphasizing that this isoquant is meaningful only prior to the construction of the machines of various types. Once a machine of a type is constructed, it cannot be substituted for one of another type.

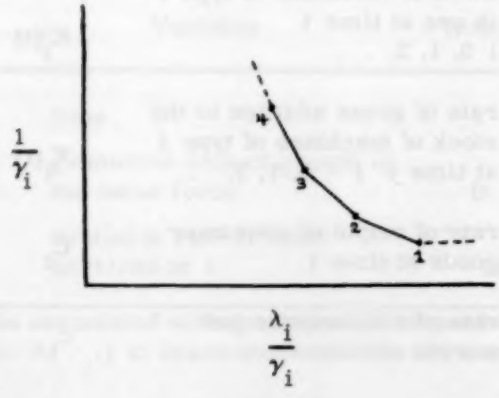


Figure 2a

1. We shall be assuming that $\lim_{i \to \infty} \gamma_i = \gamma > 0$; $\lim_{i \to \infty} \frac{\lambda_i}{\gamma_i} = \frac{\lambda}{\gamma} > 0$.

2. Labor intensity is defined as labor input per unit of output.

Consider an economy described by the technology described by (2.1)-(2.3) and a labor force growing at constant proportionate rate Θ. Suppose now that this economy is free to choose the initial stocks of various types of machines. If the planners of this economy wish to maximize the sustainable rate of consumption per worker, what should be their choices regarding the initial values and growth paths of the stocks of various types of machines?[1] Before we can answer this question, we have to define our terms.

Definition 1

A time path is a vector $c = (c_1, c_2, \ldots c_t, \ldots)$ where c_t is the consumption per worker at time t.

Definition 2[2]

A time path c is said to be feasible if we can find $K_i(0)$, $K_i'(t)$, $X_i(t)$ $(i = 0, 1, 2, \ldots)$ such that the following holds for $t = 0, 1, 2, \ldots$

$$c_{t+1} = (1+\Theta)^{-t-1} \sum_{i=1}^{\infty} \gamma_i K_i'(t) \tag{2.4}$$

$$K_i(t) = K_i(0)(1-\delta)^t + \sum_{t'=0}^{t-1} (1-\delta)^{t-1-t'} X_i(t') \qquad i = 0, 1, \ldots \tag{2.5}$$

$$K_i(t) - K_i'(t) \geq 0 \qquad i = 0, 1, 2, \ldots \tag{2.6}$$

$$-\sum_{i=0}^{\infty} \lambda_i K_i'(t) \geq -(1+\Theta)^t \tag{2.7}$$

$$\gamma_0 K_0'(t) - \sum_{i=0}^{\infty} X_i(t) \geq 0 \tag{2.8}$$

$$K_i(0),\ K_i'(t),\ X_i(t) \geq 0 \qquad i = 0, 1, 2, \ldots \tag{2.9}$$

Equation (2.4) states that consumption per worker at time (t+1) is the ratio of the output of consumer goods of the previous period to the labor force at that time.[3] Equation (2.5) states that the stock in existence at t of type i machines is equal to the depreciated level of the initial stock plus the

---

1. This particular maximand is not the only one that can be considered. This point is discussed subsequently.

2. This definition has to be modified when $K_i(0)$ $(i = 0, 1, 2, \ldots)$ are treated as historically given. Then the definition will read "A time path is said to be feasible if we can find $K_i'(t)$, $X_i(t)$ $(t = 0, 1, 2, \ldots)$ such that ...."

3. This is because of our assumption of a lag of one period in production.

depreciated level of the gross addition made at each point in time from 0 to t-1 . Inequality (2.6) ensures that the stock in use of any type of machine does not exceed the stock in existence at any t . Inequality (2.7) ensures that the rate of employment at any t does not exceed the rate of labor services available at that t . Inequality (2.8) states that gross investment at any t does not exceed the rate of output of machines at that t . Inequality (2.9) ensures that stock variables and gross investment variables are nonnegative.

Definition 3

A feasible time path c is efficient if there is no other feasible time path $\bar{c}$ such that (a) $\bar{c}_t \geq c_t$ for all t and (b) $\bar{c}_t > c_t$ for at least one t .

Definition 4

A rate of consumption per worker $\underline{c}$ is said to be sustainable if there exists a feasible time path c such that $c_t \geq \underline{c}$ for all t .

Clearly if $\underline{c}$ is sustainable, so is any $\underline{\underline{c}} < \underline{c}$. Now, from (2.7) it is clear that the stock of machines of any type in use cannot grow indefinitely at a proportionate rate greater than Θ , given the fact that $\lambda_i > 0$ for all i and $\lim_{i \to \infty} \lambda_i = \lambda > 0$. This implies that there can be no feasible time path in which consumption per worker increases without any limit. Thus in the long run the economy under consideration can sustain only a constant rate of consumption per worker. It is therefore meaningful to derive that time path which maximizes the sustainable rate of consumption per worker (MSCP) assuming that the initial stocks of various types of capital are subject to choice.

It can be shown (see Section 2.2 for proof) that, for the economy under consideration, the MSCP is associated with the time path $c^j = (c^j_1, c^j_2, \ldots)$ defined by the following:[1] for $t = 0, 1, \ldots$:

$$(1+\Theta)c^j_{t+1} = \frac{\gamma_j(\gamma_0 - \Theta - \delta)}{\lambda_j(\gamma_0 - \Theta - \delta) + \lambda_0(\Theta + \delta)} \tag{2.10}$$

$$K_0(t) = K'_0(t) = \frac{(\Theta+\delta)(1+\Theta)^t}{\lambda_j(\gamma_0 - \Theta - \delta) + \lambda_0(\Theta+\delta)} \tag{2.11}$$

---

1. It will be noticed that $\gamma_0 - \delta$ is assumed to be larger than Θ . If $\gamma_0 - \delta < \Theta$, the economy cannot sustain even a constant rate of consumption per worker. This follows from the fact that $\gamma_0 - \delta$ is the net reproduction rate of machines. If this is less than the rate of growth of labor force, consumption per worker cannot be sustained at a constant nonzero level, since by assumption there is no technique which uses only labor to produce consumer goods.

$$K_j(t) = K'_j(t) = \frac{(\gamma_0 - \Theta - \delta)(1+\Theta)^t}{\lambda_j(\gamma_0 - \Theta - \delta) + \lambda_0(\Theta+\delta)} \tag{2.12}$$

$$X_0(t) = (\Theta+\delta)K_0(t) \tag{2.13}$$

$$X_j(t) = (\Theta+\delta)K_j(t) \tag{2.14}$$

$$K_i(t) = K'_i(t) = X_i(t) = 0 \quad \text{for} \quad i \neq 0, j \tag{2.15}$$

j is an index which satisfies the following:[1]

$$\frac{\gamma_{j-1} - \gamma_j}{\lambda_{j-1}\gamma_j - \lambda_j\gamma_{j-1}} < \frac{(\gamma_0 - \Theta - \delta)}{\lambda_0(\Theta+\delta)} < \frac{\gamma_j - \gamma_{j+1}}{\lambda_j\gamma_{j+1} - \lambda_{j+1}\gamma_j} \tag{2.16}$$

The time path c^j described by (2.10) - (2.15) is a balanced growth path on which the stocks of machines of type 0 and j and the labor force grow at the same rate Θ. Consumption per worker is constant along this time path. Only one of the infinitely many alternative techniques is used for producing consumer goods. This result could be explained in two ways.

It can be seen from (2.4) - (2.8) that any feasible time path which yields a constant rate of consumption per worker is necessarily a balanced growth path. One can show that, for each $i \geq 1$, a feasible balanced growth path c^i can be obtained by replacing j by i in (2.10) - (2.15). Such a path c^i uses only technique i for producing consumer goods. Any feasible balanced growth path which uses more than one technique for producing consumer goods can be expressed as a suitable combination of the paths c^i $(i \geq 1)$. The sustainable rate of consumption per worker (SCP) of the path c^i is

$$(1+\Theta)\underline{c}^i = \frac{\gamma_i(\gamma_0 - \Theta - \delta)}{\lambda_i(\gamma_0 - \Theta - \delta) + \lambda_0(\Theta+\delta)} \tag{2.17}$$

Inequalities (2.16) together with (2.3) imply that $\underline{c}^i$ attains its maximum value when $i = j$. In other words, among the class of time paths which yield constant rates of consumption per worker over time, the path c^j

---

1. It is conceivable that for all i, $\frac{\gamma_{i-1} - \gamma_i}{\lambda_{i-1}\gamma_2 - \lambda_i\gamma_{i-1}} < \frac{(\gamma_0 - \Theta - \delta)}{\lambda_0(\Theta+\delta)}$. It can be shown that in this case the sustainable rates of consumption are bounded above, and this bound can either be attained or approached as closely as one wishes by choosing a sufficiently large j. One other possibility is that one of the inequalities in (2.16) is an equality. Only one can be an equality in view of (2.3). If this possibility occurs, the MSCP can be achieved by using either j or one of the techniques adjacent to it.

yields the maximum rate. Further, (2.16) and (2.3) also imply that if we consider two economies which differ only in their rates of growth of the labor force, the economy with the faster growing labor force will use a technique, yielding its MSCP, which is at least as labor intensive as the other economy. The MSCP of the former economy cannot exceed that of the latter. Figure 2b illustrates these results. [1]

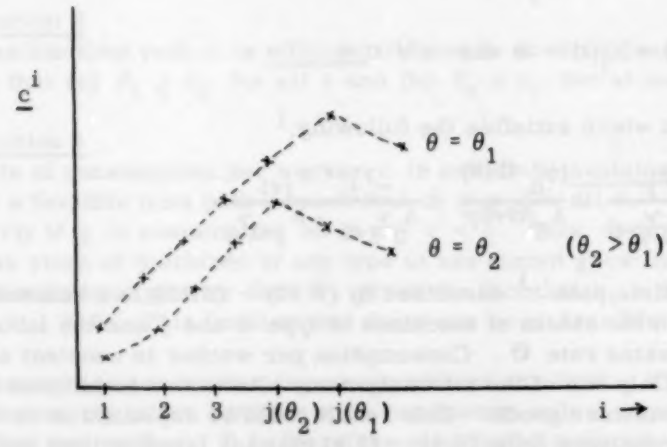


Figure 2b

Another way of interpreting the choice of j through (2.16) is the following. Suppose we wish to produce an SCP of one unit. Since the labor force at time t is $(1+\Theta)^t$, the rate of output of consumer goods at time t should then be $(1+\Theta)^t$. For producing this rate of output, one needs direct labor services of $\frac{\lambda_i}{\gamma_i}(1+\Theta)^t$ units if one uses technique i. Further, it can be shown that one also needs indirect labor services of $\left[\frac{\lambda_0(\Theta+\delta)}{\gamma_i(\gamma_0-\Theta-\delta)}\right](1+\Theta)^t$ units for producing enough machines of types 0 and i to maintain the desired rate of consumption per worker. Thus the total labor requirement [2] per unit of SCP is $\xi_i = \left[\frac{\lambda_i}{\gamma_i} + \frac{\lambda_0(\Theta+\delta)}{\gamma_i(\gamma_0-\Theta-\delta)}\right](1+\Theta)^t$. Since in the final analysis labor is the only constraint, by minimizing ξ_i by a proper choice

1. This figure is for illustrative purposes only.

2. The ratio of direct and indirect labor services needed per unit of SCP seems to correspond to Marx's concept of the organic composition of capital.

of i we can maximize the SCP. Inequalities (2.16) together with (2.3) establish that ξ_i attains its minimum value when $i = j$. In other words, as i increases, the direct labor requirement falls while the indirect labor requirement increases. Up to $i = j$, the fall in the direct labor requirement is greater than the rise in the indirect labor requirement, while beyond j the inequality is reversed and hence j is the technique which minimizes ξ_i.

In summary, it has been shown that a particular two-sector economy is capable of sustaining a certain maximum rate of consumption per worker over time. This rate is attained at each point of time along a growth path on which only one of many alternative techniques of producing consumer goods is used. In both sectors the stock of machines grows at the same rate as the labor force. Thus such a path describes a stationary state in which every relevant variable is constant, not in absolute terms, but relative to the labor force.[1] The path can therefore be described by the stock per worker of various types of machines. We shall denote the growth path derived in this section as the "terminal path" for reasons to be explained in Section 3.

2.2. Derivations

We shall show that the time path represented by (2.10) - (2.16) yields the MSCP. First we need to establish the following lemma:

Lemma:

For a sustainable rate of consumption per worker $\underline{c}$ to be the MSCP, it is sufficient that there exists an efficient path c with $c_t = \underline{c}$ for all t.

Proof:

Suppose there exists an efficient c such that $c_t = \underline{c}$ for all t. By the definition of efficiency there is no feasible c' such that

$$\begin{aligned} c'_t &\geq c_t = \underline{c} \qquad \text{for all } t \text{ and} \\ c'_t &> c_t = \underline{c} \qquad \text{for at least one } t. \end{aligned} \tag{2.18}$$

This implies that there exists no feasible c' such that $c'_t > \underline{c} + \epsilon$ where $\epsilon > 0$. In other words, no rate greater than $\underline{c}$ can be sustained. Therefore $\underline{c}$ is the MSCP.

In order to show that a feasible path is efficient we use a lemma

---

1. The "Golden Age" of Mrs. Joan Robinson describes a similar path. See Joan Robinson, The Accumulation of Capital (London: Macmillan and Company Ltd., 1958). The model described by her incorporates neutral technical progress as well.

due to Malinvaud.[1] This lemma states that a given feasible path is efficient if (a) there exists a set of nonnegative prices which sustains it and (b) the cost of inputs of period t, evaluated with such a set of prices, approaches zero as $t \to \infty$. A particular set of prices is said to sustain a given feasible path if it can be shown that the given path will result from profit maximizing behavior under the particular set of prices. The profit maximization rule can be translated into the language of activity analysis as follows: under the chosen set of prices (a) no activity makes a positive profit and (b) any activity that is operated at a positive level on the given path makes zero profit.

The basic activities of our model are the following: (we use the same symbol for an activity and its level)

1. $K_0'(t)$: the activity which produces machines at t.
2. $K_i'(t)$: the activity which produces consumer goods with technique i $i = 1, 2, \ldots$
3. $X_i(t)$: the activity which allocates a part of newly produced machines to increasing the stock of machines of type i $i = 0, 1, 2, \ldots$

Let us use the following notation for prices: (all prices are present values, present meaning time zero)

$p(t)$: price per unit of consumer good at t.
$q(t)$: price per unit of a newly produced machine at t.
$r_i(t)$: rate of rental per period of a unit of type i machine at time t.
$w(t)$: rate of wages per unit of labor at time t.

The price conditions of efficiency can be stated as follows: given a feasible solution $X_i(t)$, $K_i'(t)$, it is efficient if a set of nonnegative prices $p(t)$, $q(t)$, $r_i(t)$, $w(t)$ could be found such that the following holds:[2]

$$\left.\begin{aligned} \gamma_i\, p(t+1) &\leq \lambda_i w(t) + r_i(t) && i = 1, 2, \ldots \\ &= \lambda_i w(t) + r_i(t) && \text{if} \quad K_i'(t) > 0 \end{aligned}\right\} \quad (2.19)$$

$$\left.\begin{aligned} \gamma_0\, q(t+1) &\leq \lambda_0 w(t) + r_0(t) \\ &= \lambda_0 w(t) + r_0(t) && \text{if} \quad K_0'(t) > 0 \end{aligned}\right\} \quad (2.20)$$

1. Edmond Malinvaud, "Capital Accumulation and Efficient Allocation of Resources," Econometrica, XXI (April 1958), Lemma 5. Malinvaud's model assumes that there are only a finite number of commodities in each period. Our model has an infinite number of commodities in the form of various types of machines. We are assuming with Malinvaud that his results could be extended to this case.

2. A production lag of one period is assumed.

$$\left.\begin{aligned} q(t+1) &\geq \sum_{t'=t+1}^{\infty} (1-\delta)^{t'-t-1} r_i(t') \qquad i = 0, 1, 2, \ldots \\ &= \sum_{t'=t+1}^{\infty} (1-\delta)^{t'-t-1} r_i(t') \quad \text{if} \quad X_i(t') > 0 \end{aligned}\right\} \tag{2.21}$$

The right hand member of (2.19) represents the cost of producing γ_i units of the consumer good with technique i . If technique i is to be used, then the cost of production of a unit of consumer good should equal its price. This is stated by the second part of (2.19). (2.20) represents a similar profit maximizing condition in the production of machines. (2.19) and (2.20) together represent one-period efficiency conditions. (2.21) states intertemporal efficiency conditions. If a unit of newly produced machine is to be added to the stock of type i machines, then the present value of the returns from such an addition should equal the cost of production of the machine. This is precisely what the second part of (2.21) states.

We are now in a position to prove the efficiency of the time path c^j . The quantity solutions of c^j and the associated price solutions are: (for $t \geq 0$)

$$\left.\begin{aligned} K_0'(t) &= \frac{(\Theta+\delta)\ (1+\Theta)^t}{\lambda_j(\gamma_0-\Theta-\delta) + \lambda_0(\Theta+\delta)} \qquad X_0(t) = (\Theta+\delta)\ K'(t) \\ K_j'(t) &= \frac{(\gamma_0-\Theta-\delta)\ (1+\Theta)^t}{\lambda_j(\gamma_0-\Theta-\delta) + \lambda_0(\Theta+\delta)} \qquad X_j(t) = (\Theta+\delta)\ K_j'(t) \\ K_i'(t) &= 0 \quad i \neq 0, j \qquad\qquad X_i(t) = 0 \quad i \neq 0, j \end{aligned}\right\} \tag{2.22}$$

$$\left\{\begin{aligned} &q(t+1) = q(0)\ [1+u]^{-t-1} \quad \text{where} \quad q(0) = \frac{\lambda_0\gamma_j}{\lambda_j(\gamma_0-u-\delta) + \lambda_0(u+\delta)} \\ &w(t) = \left[\frac{\gamma_0-u-\delta}{\lambda_0}\right] q(t+1) \quad r_0(t) = r_j(t) = (u+\delta)q(t+1) \\ &p(t+1) = [1+u]^{-t-1} \qquad\qquad r_i(t) = \alpha_1 r_j(t) \quad \text{for} \quad i < j \\ &r_i(t) = \alpha_2 r_j(t) \quad \text{for} \quad i > j \quad \text{where} \quad u,\ \alpha_1,\ \alpha_2 \quad \text{are to be} \\ &\text{chosen such that} \quad \Theta < u < (\gamma_0-\delta) \qquad 0 < \alpha_1,\ \alpha_2 < 1 . \end{aligned}\right. \tag{2.23}$$

It is easy to verify that under the prices (2.23) the second parts of (2.19) – (2.21) hold. It remains to show that for $i \neq 0$, j the following

holds to show that the first parts of (2.19) – (2.21) are also satisfied.

$$\gamma_i\, p(t+1) \leq \lambda_i\, w(t) + r_i(t)$$

$$q(t+1) \geq \sum_{t+1}^{\infty} (1-\delta)^{t'-t-1}\, r_i(t') \qquad i \neq 0,\ j \tag{2.24}$$

Now from (2.23) it can be shown that $q(t+1) = \sum_{t+1}^{\alpha} (1-\delta)^{t'-t-1}\, r_j(t')$ and that $r_i(t) < r_j(t)$ for $i \neq 0, j$. Therefore the second part of (2.24) holds.

For proving the first part of (2.24), let us substitute $\frac{\lambda_j w(t)+r_j(t)}{\gamma_j}$ for $p(t+1)$. We need to show that

$$\gamma_i \frac{\lambda_j w(t)+r_j(t)}{\gamma_j} \leq \lambda_i\, w(t) + r_i(t) \quad \text{for } i \neq 0, j \tag{2.25}$$

Substituting for $r_i(t)$ from (2.23), we can reduce (2.25) to

$$\frac{\gamma_i - \alpha_1 \gamma_j}{\lambda_i\gamma_j - \lambda_j\gamma_i} \leq \frac{w(t)}{r_j(t)} = \frac{(\gamma_0 - u - \delta)}{\lambda_0(u+\delta)} \quad \text{for } i < j \tag{2.26}$$

$$\frac{\alpha_2\gamma_j - \gamma_i}{\lambda_j\gamma_i - \lambda_i\gamma_j} \leq \frac{w(t)}{r_j(t)} = \frac{(\gamma_0 - u - \delta)}{\lambda_0(u+\delta)} \quad \text{for } i > j \tag{2.27}$$

Using (2.23) it is easy to show that

$$\frac{\gamma_i - \alpha_1\gamma_j}{\lambda_i\gamma_j - \lambda_j\gamma_i} < \frac{\gamma_{j-1} - \alpha_1\gamma_j}{\lambda_{j-1}\gamma_j - \lambda_j\gamma_{j-1}} \quad \text{for } i < j-1$$

$$\frac{\alpha_2\gamma_j - \gamma_i}{\lambda_j\gamma_i - \lambda_i\gamma_j} > \frac{\alpha_2\gamma_j - \gamma_{j+1}}{\lambda_j\gamma_{j-1} - \lambda_{j+1}\gamma_j} \quad \text{for } i > j+1$$

Therefore if (2.26) and (2.27) hold for $i = j - 1$ and $i = j + 1$ respectively, they hold for all other relevant values of i. By (2.16) we know that j is an index which satisfies

$$\frac{\gamma_{j-1} - \gamma_j}{\lambda_{j-1}\gamma_j - \lambda_j\gamma_{j-1}} < \frac{(\gamma_0 - \Theta - \delta)}{\lambda_0(\Theta + \delta)} < \frac{\gamma_j - \gamma_{j+1}}{\lambda_j\gamma_{j+1} - \lambda_{j+1}\gamma_j}$$

By choosing u sufficiently close to but greater than Θ and α_1, α_2

sufficiently close to unity we can make (2.26) and (2.27) hold for $i = j - 1$ and $i = j + 1$ respectively. Thus we have succeeded in showing that we can associate a set of nonnegative prices with the path c^j that sustains it. Now we have to show that the cost of inputs of period t evaluated at the prices (2.23) goes to zero as $t \to \infty$. This is easily proved. For along the path c^j all inputs are of the form $A[1 + \Theta]^t$ and prices given by (2.23) are of the form $B[1 + u]^{-t}$. Therefore the total cost of inputs is of the form $D(1+\Theta)^t (1+u)^{-t}$, which goes to zero as $t \to \infty$ since u has been chosen to be greater than Θ.

It is interesting to note the effect of changes in Θ on the index j of the optimal technique and $\underline{c}^j$, the MSCP associated with j. It will be recalled that j is chosen so that[1]

$$\frac{\gamma_{j-1}-\gamma_j}{\lambda_{j-1}\gamma_j - \lambda_j\gamma_{j-1}} < \frac{(\gamma_0-\Theta-\delta)}{\lambda_0(\Theta+\delta)} < \frac{\gamma_j-\gamma_{j+1}}{\lambda_j\gamma_{j+1}-\lambda_{j+1}\gamma_j} \tag{2.29}$$

$$\text{and} \quad \underline{c}^j = \frac{\gamma_j(\gamma_0-\Theta-\delta)(1+\Theta)^{-1}}{\lambda_j(\gamma_0-\Theta-\delta) + \lambda_0(\Theta+\delta)} \tag{2.29}$$

It is obvious that $\dfrac{(\gamma_0-\Theta-\delta)}{\lambda_0(\Theta+\delta)}$ is a decreasing function of Θ. From (2.23) we know that

$$\frac{\gamma_{i-1}-\gamma_i}{(\lambda_{i-1}\gamma_i - \lambda_i\gamma_{i-1})} < \frac{\gamma_i - \gamma_{i+1}}{(\lambda_i\gamma_{i+1} - \lambda_{i+1}\gamma_i)} \quad \text{for all } i .$$

Therefore as Θ increases the optimal j can never increase. This also implies that optimal $\underline{c}^j$ decreases as Θ increases. For, given any j, $\underline{c}^j$ decreases as Θ increases.

3. The Approach to the Terminal Path

3.1. Results

In Section 2 the "terminal path" was defined to be a path which maximized the sustainable rate of consumption per worker. It was assumed that even the initial stocks of machines of various types were subject to choice, and this enabled us to initiate the economy on the terminal path itself. But in most development problems it is more realistic to assume that initial conditions, such as the stocks of machines of various types, are

---

1. It will be assumed that the j which satisfies (2.28) is such that $\lambda_j \geq \lambda_0$. If $\lambda_0 < \lambda_j$ it can be shown that feasible paths exist which converge to the path c^j from above. Thus the path which yields the MSCP is not unique.

historically given. These initial stocks will not normally be equal to their corresponding values on the terminal path. Then we will not be able to start the economy on the terminal path.

The choices open to such an economy could be analyzed in a formal way as follows: define a feasible path c, as in Section 2, modifying it to take into account that we can choose only $K_i(t)$ for all t but not $K_i(0)$.[1] Let us postulate that social preferences could be represented in terms of an ordinal utility function $U(c)$.[2] The objective of the planners, then, is to choose a feasible path c which maximizes $U(c)$.

Such a formulation is much too general. In the present section we consider one particular utility indicator. We restrict ourselves to the subset S of the set of all feasible paths, which consists of time paths for which the following is true: $c_t = c^E$ for all $t \geq$ some finite T (depending on the particular path c), where c^E is the rate of consumption per worker associated with the terminal path. The objective will then be to choose a $c \in S$ such that the associated T is minimized. In other words, starting from given initial conditions, we wish to choose that feasible time path which reaches the terminal path in the minimum time and remains on it thereafter.

This objective completely disregards the time path of the rate of consumption per worker prior to the time of arrival at the terminal path. A somewhat more realistic approach will be to minimize the time of arrival T at the terminal path subject to the constraint that the rate of consumption per worker never falls below a certain floor at any time $t < T$. This approach is not followed in this section for reasons of analytical convenience.

Even with the simplifying assumption of disregarding consumption prior to T, we have not been able to solve the problem of choosing a time path $c \in S$ which minimizes T. However, we have been able to solve for a time path c which minimizes T among the class of time paths which are members of a subset S' of S. The character of the solution obtained leads us to the conjecture that the solution does, in fact, minimize T among all $c \in S$ as well.

The particular subset S' will consist of feasible paths c with the following characteristics: (1) each such path reaches the terminal path at a time $T(\sigma)$;[3] (2) during an interval $[0, t_1(\sigma)]$ of time $(t_1(\sigma) < T(\sigma))$, the stock of machines of type 0 is made to grow at a rate $\sigma > \Theta$ from an initial level below its value on the terminal path; (3) at any t in $0 \leq t \leq t_1(\sigma)$,

---

1. See footnote 2, Section 2, p. 67.

2. We are <u>not</u> starting from individual preference orderings to obtain a social preference ordering. By postulating a social preference ordering itself, we assume away the problem of relating social preference ordering to individual preference orderings.

3. In this section time will be treated as a continuous variable.

machine production in excess of that needed to assure a growth rate σ of type 0 machines is invested in type j machines;[1] (4) at $t_1(\sigma)$ the stock of type 0 machines is sufficiently above its corresponding value on the terminal path that we can attain the terminal path at $T(\sigma)$ by letting it depreciate; and (5) the entire output of machines at any t in $t_1(\sigma) < t \leq T(\sigma)$ is invested in type j machines. Each path $c \in S'$ can therefore be defined by its σ. The range for σ is the semi-open interval $(\Theta, \gamma_0-\delta]$.

Unless σ is greater than the rate of growth of the labor force, Θ, the stock of machines of type 0 cannot grow from an initial level below that corresponding to the terminal path to a level above the terminal path at $t_1(\sigma)$. The rate of growth of the stock of machines (and hence σ) cannot exceed $(\gamma_0-\delta)$, the maximum net reproduction rate of machines.

The path which minimizes $T(\sigma)$, the time of arriving at the terminal path, is one for which $\sigma = \gamma_0 - \delta$. This means that during an interval $[0, t_1(\gamma_0 - \delta)]$ of time, the stock of machines of type 0 grows at the maximum feasible rate $(\gamma_0-\delta)$ and, necessarily, the existing stocks of machines of all other types decline at the maximum possible rate δ. Hence consumption per worker declines during the same interval of time. In algebraic terms the path which minimizes $T(\sigma)$ is given by the following equations:[2]

$$\left.\begin{aligned} K_0(t) &= K_0(0)\, e^{(\gamma_0-\delta)t} && (3.1)\\ K_j(t) &= K_j(0)\, e^{-\delta t} && (3.2) \end{aligned}\right\} \quad 0 \leq t \leq t_1(\gamma_0-\delta)$$

$$\left.\begin{aligned} K_0(t) &= K_0(0)\, e^{\left[(\gamma_0-\delta)t_1 - \delta(t-t_1)\right]} && (3.3)\\ K_j(t) &= \left[\gamma_0(t-t_1)\, K_0(0)\, e^{(\gamma_0-\delta)t_1} + K_j(0) e^{-\delta t_1}\right] e^{-\delta(t-t_1)} && (3.4) \end{aligned}\right\} \quad t_1 \leq t \leq T$$

where $t_1(\gamma_0-\delta)$ and $T(\gamma_0-\delta)$ are to be solved from

---

1. It is worth recalling that on the terminal path the stocks of machines 0 and j only are different from zero and grow at the rate Θ.

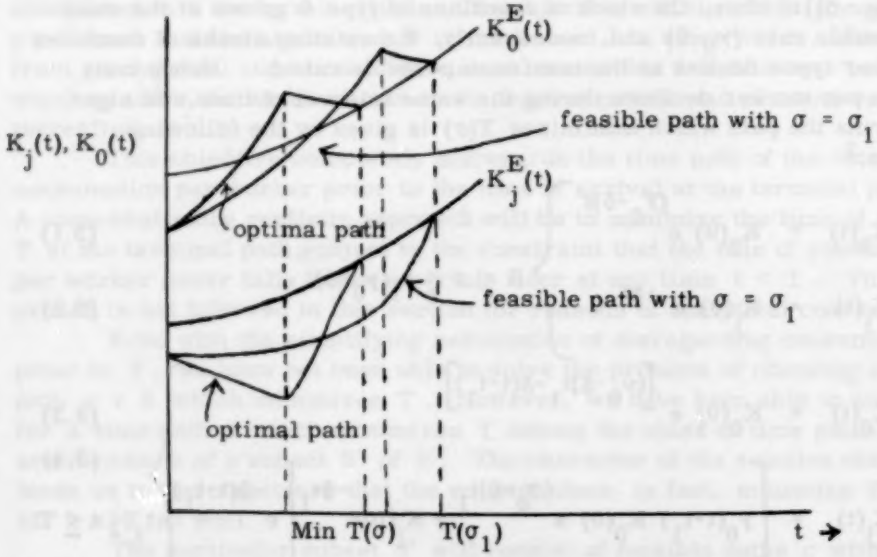
2. For brevity, only the stocks (in existence) of machines of type 0 and j are given explicitly. The given initial stocks of all other types of machines are allowed to depreciate along this path. In (3.3) – (3.6), the argument $(\gamma_0-\delta)$ associated with $t_1(\gamma_0-\delta)$ and $T(\gamma_0-\delta)$ is omitted to avoid confusion. A superscript E is used to denote terminal path values.

$$K_0(T) = K_0^E(T) = \frac{(\Theta+\delta)e^{\Theta T}}{\lambda_j(\gamma_0-\Theta-\delta) + \lambda_0(\Theta+\delta)} \tag{3.5}$$

$$K_j(T) = K_j^E(T) = \frac{(\gamma_0-\Theta-\delta)e^{\Theta T}}{\lambda_j(\gamma_0-\Theta-\delta) + \lambda_0(\Theta+\delta)} \tag{3.6}$$

We can illustrate two members of S', one of which minimizes $T(\sigma)$, in the following figure:[1]

Figure 3a



The trade-off diagram below shows that as σ increases from Θ maximum possible value $\gamma_0-\delta$, $T(\sigma)$ declines continuously. The larger the value of σ, the greater is the cost (loosely speaking), in terms of consumption foregone prior to $T(\sigma)$, of reaching the terminal path. This can be seen as follows. Suppose the planners have made a decision to reach the terminal path through a member of the class of paths S'.[2]

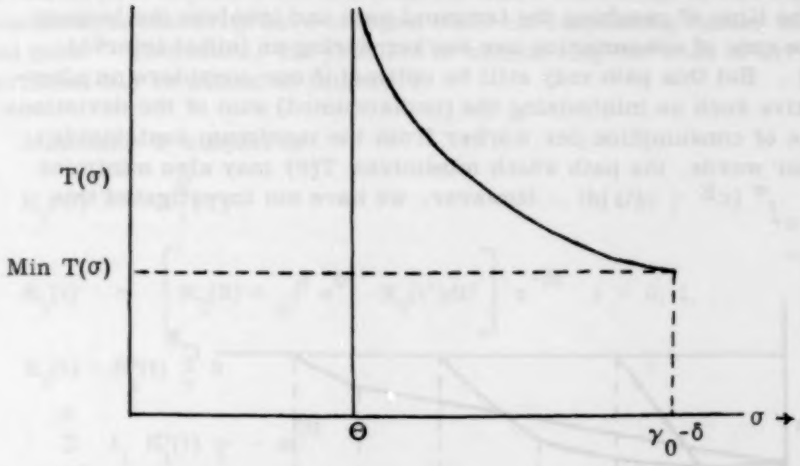
---

1. This and all subsequent figures are for illustrative purposes only. No quantitative significance is intended.

2. A complete discussion of the cost, in terms of consumption foregone (suitably defined), of reaching the terminal path in a finite time will

Then their choice is restricted to some σ in $\Theta < \sigma \leqq (\gamma_0 - \delta)$.

Figure 3b



It can be shown (see Section 2) that along any time path $c \in S'$ the rate of labor services devoted to the production of machines at any t in $0 \leqq t \leqq T(\sigma)$ is less than the available supply. Thus a surplus of labor is available at each t for employment in the production of consumer goods. Further, by definition of S', σ is the rate of growth of the stock of machines of type 0 (machine producing machines) at any t in $0 \leqq t \leqq t_1(\sigma) < T(\sigma)$. Hence the larger the value of σ, the smaller is the surplus of labor available for employment in the production of consumer goods, and also the smaller is the stock of various types of machines capable of producing consumer goods. Because of both these factors, the possible rate of production of consumer goods is smaller (at any t in $0 \leqq t \leqq t_1(\sigma)$) the larger the value of σ. It can be shown that $\sigma = \gamma_0 - \delta$ minimizes both $t_1(\sigma)$ and $T(\sigma)$. Hence the path associated with $\sigma = (\gamma_0 - \delta)$ involves the maximum sacrifice in terms of consumption foregone during $0 \leqq t \leqq \underset{\sigma}{\text{Min}}\, t_1(\sigma) = t_1(\gamma_0 - \delta)$. We can illustrate this in terms of the following figure.

In Figure 3c, three possible time paths of consumption per worker are shown corresponding to three different values of σ in the order $\gamma_0 - \delta > \sigma_1 > \sigma_2$. The figure as drawn shows a path which permits

---

involve comparing feasible paths $c \in S$. We are again restricting ourselves to paths $c \in S' \subset S$ by assuming that the planners have made an arbitrary decision to choose a $c \in S'$. Since the discussion above is of a qualitative character, this assumption is not very serious.

consumption per worker to increase continuously during $0 \leqq t \leqq T(\sigma)$. In general, such a path need not exist. For if the initial stock of machines of type 0 is not sufficiently large, any path, whether it leads to the terminal path or not, will involve a decline in the rate of consumption per worker during an initial interval of time.[1] In any case, it is clear that the path minimizes the time of reaching the terminal path and involves the largest decline in the rate of consumption per worker during an initial interval $[0\ ,\ t_1(\gamma-\delta)]$. But this path may still be optimal if one considers an alternative objective such as minimizing the (undiscounted) sum of the deviations of actual rate of consumption per worker from the maximum sustainable rate. In other words, the path which minimizes $T(\sigma)$ may also minimize the integral $\int_0^\infty [c^E - c(t)]dt$. However, we have not investigated this approach.

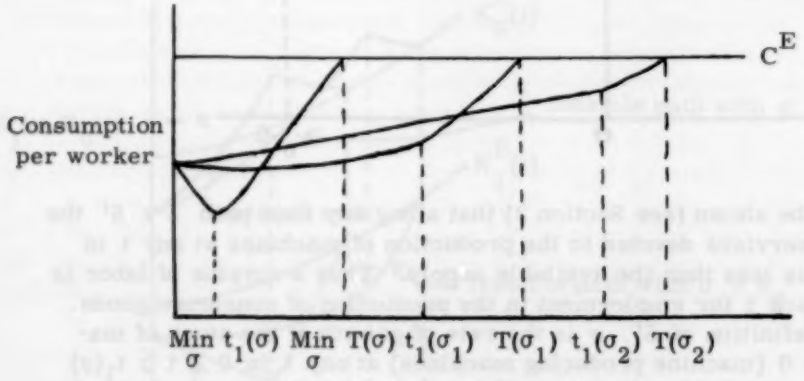


Figure 3c

1. This assertion can be proved as follows: Suppose the initial rate of output of consumer goods is $C(0)$. If the initial consumption per worker is to be maintained, the rate of output of consumer goods has to increase at the rate $\Theta C(0)$. The rate of gross increase in machines needed to achieve this is $\frac{(\Theta+\delta)C(0)}{\gamma_1}$ if we use technique 1, the least machine intensive of the alternative techniques of producing consumer goods. On the other hand, the maximum rate of output of machines is $\gamma_0 K_0(0)$ where $K_0(0)$ is the initial stock of machines of type 0 . Thus, if $\gamma_0 K_0(0) < \frac{(\Theta+\delta)C(0)}{\gamma_1}$, any time path will involve a decline in the rate of consumption per worker at time 0 .

3.2. Derivations

We shall derive the path which reaches the terminal path from given initial conditions in minimum time and thereafter coincides with the terminal path. It can be easily shown that along any path which coincides with the terminal path beyond a certain point in time, the stocks of various types of machines necessarily have to equal their corresponding values on the terminal path.[1] Therefore, the problem of minimizing the time of arrival at the terminal can be stated as follows:

Minimize T subject to

$$K_i(T) = K_i^E(T) \qquad i = 0, j \qquad (3.7)$$

$$K_i(t) = \left[K_i(0) + \int_0^t e^{\delta t'} X_i(t')dt' \right] e^{-\delta t} \qquad i = 0, 1, \ldots \qquad (3.8)$$

$$K_i(t) - K_i'(t) \geq 0 \qquad (3.9)$$

$$-\sum_{i=0}^{\infty} \lambda_i K_i'(t) \geq -e^{\Theta t} \qquad (3.10)$$

$$\gamma_0 K_0'(t) - \sum_{i=0}^{\infty} X_i(t) \geq 0 \qquad (3.11)$$

$$K_i'(t), \; X_i(t) \geq 0 \qquad (3.12)$$

$$0 \leq t \leq T$$

Restraint (3.7) states that the terminal path is attained at T. Equation (3.8) gives the stock in existence of type j machine at any time t. Restraint (3.9) ensures that the stock in use of any type of machine does not exceed the stock in existence. (3.10) is the labor restraint. Restraint (3.11) ensures that gross investment does not exceed the rate of output of new machines. (3.12) is the nonnegativity restraint.

It is clear that if one wanted to minimize T subject to the additional constraint that the rate of consumption per worker at any $t < T$ never fell below a certain floor $\bar{c}$, then one must add the following to the set of restraints (3.7) – (3.12)

$$\sum_{i=1}^{\infty} \gamma_i K_i'(t) \geq \bar{c} e^{\Theta t} .$$

However, we shall not follow this approach.

---

1. This result is true under the assumption that $\lambda_0 \leq \lambda_j$.

It was mentioned in Section 3.1 that T is to be minimized among paths belonging to a particular subset S' of the class of paths satisfying (3.7) – (3.12). We can express a member of S' in algebraic terms as follows:[1]

for t in $0 \leq t \leq t(\sigma) = t_1$ define

$$K_0(t) = K_0'(t) = K_0(0)e^{\sigma t} \tag{3.13}$$

$$K_j(t) = \left(\frac{\gamma_0 - \sigma - \delta}{\sigma + \delta}\right) K_0(0)e^{\sigma t} + \left\{K_j(0) - \left(\frac{\gamma_0 - \sigma - \delta}{\sigma + \delta}\right) K_0(0)\right\} e^{-\delta t} \tag{3.14}$$

$$K_i(t) = K_i(0)e^{-\delta t} \qquad \text{for } i \neq 0, j \tag{3.15}$$

$$K_i'(t) = 0 \qquad \text{for } i \neq 0 \tag{3.16}$$

$$X_0(t) = (\sigma+\delta)\ K_0(0)e^{\sigma t} \tag{3.17}$$

$$X_j(t) = (\gamma_0 - \sigma - \delta)\ K_j(0)e^{\sigma t} \tag{3.18}$$

$$X_i(t) = 0 \qquad \text{for } i \neq 0, j \tag{3.19}$$

and for t in $t(\sigma) < t \leq T(\sigma) = t_2$ define

$$K_0'(t) = K_0(t) = K_0(t_1)e^{-\delta(t-t_1)} \tag{3.20}$$

$$K_j(t) = [K_j(t_1) + \gamma_0(t-t_1)\ K_0(t_1)]e^{-\delta(t-t_1)} \tag{3.21}$$

$$K_i(t) = K_i(0)e^{-\delta t} \qquad \text{for } i \neq 0, j \tag{3.22}$$

$$K_i'(t) = 0 \text{ for } i \neq 0 \tag{3.23}$$

$$X_j(t) = \gamma_0\ K_0(t_1)e^{-\delta(t-t_1)} \tag{3.24}$$

$$X_i(t) = 0 \qquad \text{for } i \neq j \tag{3.25}$$

---

1. It is convenient to write t_1 for $t(\sigma)$ and t_2 for $T(\sigma)$ to avoid confusion.

t_1 and t_2 are determined from

$$K_0'(t_2) = K_0^E(t_2) \tag{3.26}$$

$$K_j'(t_2) = K_j^E(t_2) \tag{3.27}$$

By varying σ in the semi-open interval $\Theta < \sigma \leqq (\gamma_0 - \delta)$ we can get all the members of S'. We have not yet shown that S' is a subset of the class of paths satisfying (3.7) – (3.12). Assuming that for a given σ we can find $t_2 \geqq t_1 > 0$ such that (3.26) and (3.27) hold, we can verify that each member of S' satisfies all the relevant restraints except (3.10). We shall prove that (3.10) is also satisfied after proving that t_1 and t_2 exist.

After substituting for $K_0'(t_2)$, $K_j'(t_2)$, $K_0^E(t_2)$ and $K_j^E(t_2)$, (3.26) and (3.27) can be rewritten as

$$K_0(0)e^{\sigma t_1 - \delta(t_2 - t_1)} = K_0^E(0)e^{\Theta t_2} \tag{3.28}$$

$$\left[\left\{\left(\frac{\gamma_0 - \sigma - \delta}{\sigma + \delta}\right) + \gamma_0(t_2 - t_1)\right\} K_0(0)e^{\sigma t_1} + \left\{K_j(0) - \left(\frac{\gamma_0 - \sigma - \delta}{\sigma + \delta}\right) K_0(0)\right\} e^{-\delta t_1}\right] e^{-\delta(t_2 - t_1)} = K_j^E(0)e^{\Theta t_2} \tag{3.29}$$

$$\text{Define } \sigma_0 + \delta = \frac{\gamma_0 K_0(0)}{K_0(0) + K_j(0)} \tag{3.30}$$

Now $\dfrac{K_0^E(0)}{K_j^E(0)} = \dfrac{(\Theta + \delta)}{(\gamma_0 - \Theta - \delta)}$ by definition of $K_0^E(0)$ and $K_j^E(0)$.

Eliminating t_2 between (3.28) and (3.29) and using (3.30) we get

$$\left(\frac{\sigma - \Theta}{\Theta + \delta}\right)t_1 + \left[\frac{1}{\sigma_0 + \delta} - \frac{1}{\sigma + \delta}\right] e^{-(\sigma+\delta)t_1} = \left(\frac{1}{\Theta + \delta} - \frac{1}{\sigma + \delta}\right) + \frac{1}{(\Theta + \delta)} \log\left(\frac{K_0^E(0)}{K_0(0)}\right) \tag{3.31}$$

Let us define

$$f(t) = \left(\frac{\sigma - \Theta}{\Theta + \sigma}\right)t + \left[\frac{1}{\sigma_0 + \delta} - \frac{1}{\sigma + \delta}\right]e^{-(\sigma+\delta)t} \tag{3.32}$$

and

$$k = \left(\frac{1}{\Theta + \delta} - \frac{1}{\sigma + \delta}\right) + \frac{1}{(\Theta + \delta)} \log\left(\frac{K_0^E(0)}{K_0(0)}\right) \tag{3.33}$$

We wish to examine whether a nonnegative solution to the equation $f(t) = k$ exists. The discussion is facilitated by distinguishing a few cases:

Case 1. $\sigma_0 \geq \Theta$

It is obvious that[1] $f(0) < k$ given that $K_0^E(0) > K_0(0)$. Now

$$\dot{f}(t) = \frac{df}{dt} = \left(\frac{\sigma - \Theta}{\Theta + \delta}\right) - \left(\frac{\sigma - \sigma_0}{\sigma_0 + \delta}\right)e^{-(\sigma+\delta)t} \tag{3.34}$$

Clearly $\dot{f}(t) > 0$ for all $t \geq 0$ since $\sigma_0 \geq \Theta$ and $\sigma > \Theta$. It is also clear that $f(t) \to \infty$ as $t \to \infty$. Therefore a unique $t_1 > 0$ exists such that $f(t_1) = k$.

Case 2. $\sigma_0 < \Theta$

It is clear from (3.34) that $\dot{f}(t) < 0$ for $0 \leq t < t'$ and $\dot{f}(t) > 0$ for $t > t' > 0$ where t' is given by

$$\dot{f}(t') = 0 \tag{3.35}$$

One can verify, by examining the second derivative of $f(t)$ at $t = t'$, that $f(t)$ attains its minimum value at $t = t'$. Now it can be shown that $f(t') < k$.[2] Therefore there exists a unique $t_1 > t' > 0$ such that $f(t_1) = k$.

We now show that $t_2 > t_1$. From (3.28) we have

$$(t_2 - t_1) = \left(\frac{\sigma - \Theta}{\Theta + \delta}\right)t_1 - \left(\frac{1}{\Theta + \delta}\right)\log\left(\frac{K_0^E(0)}{K_0(0)}\right) \tag{3.36}$$

---

1. It is worth remembering that approaches to the terminal from below are being considered. In other words, we are starting from $K_0(0) < K_0^E(0)$ and $K_j(0) < K_j^E(0)$.

2. The proof of this assertion is rather long and tedious. Since it does not yield any results that are used subsequently, it has been omitted.

Using (3.31) with (3.36) we get

$$(t_2 - t_1) = \left(\frac{1}{\Theta+\delta} - \frac{1}{\sigma+\delta}\right) - \left(\frac{1}{\sigma_0+\delta} - \frac{1}{\sigma+\delta}\right) e^{-(\sigma+\delta)t_1} \tag{3.37}$$

$$= \frac{1}{(\sigma+\delta)} \dot{f}(t_1) \tag{3.38}$$

We know that if $\sigma_0 \geq \Theta$, $\dot{f}(t) \geq 0$ for all $t \geq 0$ and hence $(t_2 - t_1) > 0$. If $\sigma_0 < \Theta$, we know that $t_1 > t'$ and $\dot{f}(t) > 0$ for $t > t'$. Therefore, in either case $(t_2 - t_1) > 0$. We have thus succeeded in showing the existence of $t_2 > t_1 > 0$ satisfying (3.26) and (3.27).

We still have to show that (3.10) is satisfied by each member of the set S' of paths described by (3.13) – (3.27). It can be seen that for any number S', if (3.10) holds at $t = t_1$, then it holds for all t in $0 < t < t_1$ and $t_1 < t \leq t_2$. Therefore we have to show that

$$\lambda_0 K_0(0) e^{\sigma t_1} \leq e^{\Theta t_1} \tag{3.39}$$

or

$$(\sigma - \Theta)t_1 + \log(\lambda_0 K_0(0)) \leq 0 \tag{3.40}$$

Substituting for $(\sigma - \Theta)t_1$ from (3.31) we get

$$(\sigma-\Theta)t_1 + \log(\lambda_0 K_0(0)) = \log\left\{\lambda_0 K_0^E(0)\right\} + \left(\frac{\sigma-\Theta}{\sigma+\delta}\right) - \left[\frac{(\sigma-\sigma_0)(\Theta+\delta)}{(\sigma+\delta)(\sigma_0+\delta)}\right] e^{-(\sigma+\delta)t_1} \tag{3.41}$$

Now

$$\left.\begin{aligned} &\left(\frac{\sigma-\Theta}{\sigma+\delta}\right) - \left[\frac{(\sigma-\sigma_0)(\Theta+\delta)}{(\sigma+\delta)(\sigma_0+\delta)}\right] e^{-(\sigma+\delta)t_1} < \left(\frac{\sigma-\Theta}{\sigma+\delta}\right) \\ &\quad \text{if } \sigma > \text{Max}(\sigma_0, \Theta) \\ &\leq \left(\frac{\sigma_0-\Theta}{\sigma_0+\delta}\right) \text{if } \sigma_0 \geq \sigma > \Theta \end{aligned}\right\} \tag{3.42}$$

By definition, $\sigma_0 < (\gamma_0 - \delta)$ and $\sigma \leq (\gamma_0 - \delta)$; therefore we can assert that

$$(\sigma-\Theta)t_1 + \log\left(\lambda_0 K_0(0)\right) \leq \frac{(\gamma_0-\Theta-\delta)}{\gamma_0} + \log\left\{\lambda_0 K_0^E(0)\right\}$$

Given that[1] $\lambda_j > \lambda_0$ it can be shown that

$$\left(\frac{\gamma_0 - \Theta - \delta}{\gamma_0} \right) + \log \left(\lambda_0 K_0^E(0) \right) \leq 0 . \qquad (3.43)$$

This completes the proof that each member of the set S' of paths described by (3.13) – (3.27) is feasible in the sense of satisfying (3.7) – (3.12). We can now show that $t_2(\sigma)$ decreases as σ increases and therefore that the member of S' which minimizes $t_2(\sigma)$ for σ in $\Theta < \sigma \leq (\gamma_0 - \delta)$ is obtained by setting $\sigma = (\gamma_0 - \delta)$.

By (3.36) $t_2 = \left(\frac{\sigma + \delta}{\Theta + \delta} \right) t_1 - \frac{1}{(\Theta+\delta)} \log \left(\frac{K_0^E(0)}{K_0(0)} \right)$

Therefore

$$\frac{dt_2}{d\sigma} = \left(\frac{1}{\Theta+\delta} \right) \left\{ (\sigma+\delta) \frac{dt_1}{d\sigma} + t_1 \right\}$$

$$= \frac{1}{(\Theta+\delta)} \frac{d\eta}{d\sigma} \quad \text{where } \eta = (\sigma+\delta)t_1$$

We can rewrite (3.31) as

$$\left(\frac{1}{\Theta+\delta} - \frac{1}{\sigma+\delta} \right) \eta + \left(\frac{1}{\sigma_0+\delta} - \frac{1}{\sigma+\delta} \right) e^{-\eta} = k \qquad (3.44)$$

Differentiating both sides of (3.44) with respect to σ we have

$$\frac{1}{(\sigma+\delta)^2} \cdot \eta + \left(\frac{1}{\Theta+\delta} - \frac{1}{\sigma+\delta} \right) \frac{d\eta}{d\sigma} - \left(\frac{1}{\sigma_0+\delta} - \frac{1}{\sigma+\delta} \right) e^{-\eta} \cdot \frac{d\eta}{d\sigma}$$

$$+ \frac{1}{(\sigma+\delta)^2} e^{-\eta} = \frac{dk}{d\sigma} \qquad (3.45)$$

Now $\frac{dk}{d\sigma} = \frac{1}{(\sigma+\delta)^2}$. Therefore (3.45) reduces to

$$\frac{d\eta}{d\sigma} \left\{ \left(\frac{1}{\Theta+\delta} - \frac{1}{\sigma+\delta} \right) - \left(\frac{1}{\sigma_0+\delta} - \frac{1}{\sigma+\delta} \right) e^{-\eta} \right\}$$

$$= \frac{1}{(\sigma+\delta)^2} (1 - e^{-\eta} - \eta) \qquad (3.46)$$

---

1. See footnote on page 75.

Using (3.37) we can rewrite (3.46) as

$$\frac{d\eta}{d\sigma} = \frac{(1 - e^{-\eta} - \eta)}{(\sigma+\delta)^2 (t_2 - t_1)} \tag{3.47}$$

Now $(1 - e^{-\eta} - \eta) < 0$ for $\eta > 0$. We know that $(t_2 - t_1) > 0$. Therefore $\frac{d\eta}{d\sigma} < 0$, implying that $\frac{dt_2}{d\sigma}$ as well as $\frac{dt_1}{d\sigma}$ are negative. Hence our assertion that $\sigma = (\gamma_0 - \delta)$ yields the path which minimizes $t_2(\sigma)$ for σ in $\Theta < \sigma \leqq (\gamma_0 - \delta)$ is proved.

4. Social Welfare as a Function of Consumption at a Point in Time

4.1 Results

One of the many social welfare functions that have been proposed in the literature makes social welfare an increasing function of the rate of consumption per worker at some chosen future point in time, say T.[1, 2] Maximizing social welfare is then equivalent to maximizing the rate of consumption per worker at T. In the present section some consequences of following this objective are discussed.

The objective of maximizing the rate of consumption at any single point of time T in the future is rather difficult to justify. It implies that no consideration is to be given to the rate of consumption at any time before or after T. However, under the following somewhat restrictive assumptions one can justify such an objective. Suppose we consider a one commodity economy, i.e., an economy in which a single commodity, say "Shmoo," serves both as a consumer good and as a capital good.[3] In other words, "Shmoos" can be either consumed or accumulated as capital stock. Suppose, further, that each unit of the stock of Shmoos at any period of time makes available a flow of (say) ρ Shmoos during that period of time. If the total human population and its rate of growth are exogenous factors, then maximizing the rate of consumption per head during any period of time T is equivalent to maximizing the stock of Shmoos at T. A necessary condition

---

1. In this section time is treated as a discrete variable.

2. Galenson and Leibenstein, op. cit.

3. Shmoo is an imaginary animal introduced to the world by Alfred Caplin. Shmoos "lays eggs at th' slightest excuse!! They also gives milk!! And as fo' meat, broiled they make the finest steaks—fried they come out the yummiest chicken—And as for upkeep they don't eat anything—there is no waste!! The hide makes the finest leather—or cloth, depending on how thick you slice it!" See Alfred Caplin (Al Capp), "The Life and Times of the Shmoo (New York: Simon and Shuster Inc., 1948), pp. 14–15.

for the latter is that the stock of Shmoos be maximized at all periods before T as well. This implies that if the target date T is moved to $T' > T$, the optimal path of the stock of Shmoos (i.e., that path which maximizes the stock of Shmoos at T') up to the period T remains unaltered. This insensitivity of the optimal path to the choice of T makes the objective of maximizing the rate of consumption per head at some period T in the future not unreasonable. This is especially true if one imposes a floor on the rate of consumption per head at all periods before T.

Once the single commodity assumption of the previous paragraph is relaxed, the objective of maximizing the rate of consumption per worker at some T loses it appeal. In an economy with two sectors, one of which produces consumer goods and the other machines, maximizing the capacity (in terms of the stocks of various types of machines) of the sector producing consumer goods at all periods before T is not a necessary condition for maximizing its capacity at T. The optimal path of the capacity of the consumer goods sector is sensitive to the choice of T.

The reason for this sensitivity can be seen as follows. In a two sector economy there are two alternative investment opportunities in each period of time as opposed to only one in a Shmoo economy. One can either increase the stock of machines in Sector I, which produces machines, or in Sector II, which produces consumer goods. These two investment opportunities make available different streams of output of consumer goods. In the notation of Section 2, an increment of one machine to the stock of type j machines in Sector II in period (T - 1) makes available an additional (maximal) flow of $\gamma_j[1 - \delta]^{t-T}$ units of consumer goods in each $t \geq T$. But an increment of one unit to the stock of type 0 machines (i.e., the machines that produce machines) in period (T - 1) increases the flow of consumer goods by $\gamma_j[1 + \gamma_0 - \delta](1 - \delta)^{t-T-1}$ in each $t \geq T + 1$, assuming that the increased capacity of Sector II is used to produce machines of type j in each period.[1] Thus by waiting for one period one can get an augmented flow of consumer goods.

The decision as to which of the two investment alternatives described above is to be chosen at each period prior to T depends on the choice of T. In the rest of this section we examine the time paths prior to T of the variables consumption per worker and capacity of Sector II that result from alternative values of T. We conclude from this examination that the objective considered, viz., maximizing the rate consumption per worker at time T, is not desirable in a two sector economy.

Since the initial stocks of various types of machines are historically given, it is enough if we devote our attention to describing the time paths of the addition to the stock of machines of various types. It can be shown (see Section 4.2) that along a path which maximizes the rate of

1. A production lag of one period is assumed in the production of machines as well as consumer goods. This is not essential to the argument.

consumption per worker at time T, the gross addition to the stock of machines in Sector II at any $t < T$ will be divided among alternative types in constant proportions independent of t.[1] Hence it is enough to describe the time paths of the following two components of gross investment: (a) $X_0(t)$, the gross addition to the stock of machines (of type 0) in Sector I, and (b) $X(t)$, the gross addition to the stock of machines in Sector II without specifying the manner in which $X(t)$ is to be distributed among alternative types.

The solutions for $X_0(t)$ and $X(t)$ fall under three cases depending on the value of T. In order to distinguish the three cases we need to define two nonnegative integers m and M such that:

$$\gamma_0 m \leqq (1 - \delta) \quad \text{and} \quad \gamma_0(m + 1) > (1 - \delta) \tag{4.1}$$

$$\lambda_0 K_0(0)\,[1+\gamma_0-\delta]^M < (1+\Theta)^M \quad \text{and} \quad \lambda_0 K_0(0)\,[1+\gamma_0-\delta]^{M+1} \geqq (1+\Theta)^{M+1} \tag{4.2}$$

From (4.2) it is clear that (M+1) is the number of periods needed to accumulate enough machines of type 0 to employ the entire labor force in Sector I. We can now enumerate the solutions for $X_0(t)$ and $X(t)$.

Case 1. $T \leqq (m+1)$. The solution is[2]

$$\left.\begin{aligned} X(t) &= \gamma_0 K_0(0)\,[1 - \delta]^t \\ X_0(t) &= 0 \end{aligned}\right\} t = 0, 1, 2, \ldots, T-2 \tag{4.3}$$

The equations of (4.3) imply that during the time span 0 - (T - 2), positive gross investment is made only in the consumer goods industry. In other words, if the time point T at which the rate of consumption per worker is to be maximized is not too far in the future, the optimal policy is to devote the entire output of Sector I to increasing the stock of machines in Sector II during the periods 0 - (T-2). Thus by letting the

---

1. This assertion will be true only when there is no floor imposed on the rate of consumption per worker at any $t < T$.

2. The range for the time variable is from period 0 to period T - 2. This follows from our assumption of a lag of one period in the production of both machines and consumer goods. The rate of consumption at T depends on the stock of machines in Sector II at T - 1, and this, in turn, depends on the investment policy from period 0 to period T - 2.

stock of machines in Sector I decline, consumption at some point of time in the near future is maximized at the expense of consumption in the far future.

Case 2a. $T > (m+1)$ and $T - m - 2 \leq M$.

The solution in this case is

$$\left.\begin{aligned} X_0(t) &= \gamma_0 K_0(0)\ [1+\gamma_0-\delta]^t \\ X(t) &= 0 \end{aligned}\right\} \quad t = 0, \ldots T-m-3 \qquad (4.4)$$

$$\left.\begin{aligned} X_0(t) &= 0 \\ X(t) &= \gamma_0 K_0(0)\ [1+\gamma_0-\delta]^{T-m-2}\ (1-\delta)^{t-(T-m-2)} \end{aligned}\right\} \quad t = T-m-2, \ldots T-2 \qquad (4.5)$$

The implication of this solution is that, for values of T in the range considered, the optimal investment policy is to devote the entire output of machines to increasing the stock of machines in Sector I during the periods 0 to (T-m-3) and to increasing the stock of machines in Sector II during the subsequent periods (T-m-2) to (T-2). Thus the rate of consumption per worker decreases during the periods 0 to (T-m-3).

Case 2b. $T > (m+1)$ and $(T-m-2) > M$. The solution in this case is:

$$\left.\begin{aligned} X_0(t) &= \gamma_0 K_0(0)\ [1+\gamma_0-\delta]^t \\ X(t) &= 0 \end{aligned}\right\} \quad t = 0, 1, \ldots M-1 \qquad (4.6)$$

$$\left.\begin{aligned} X_0(t) &= \frac{1}{\lambda_0}\ [1+\Theta]^{M+1} - (1-\delta)\ K_0(0)\ [1+\gamma_0-\delta]^M \\ X(t) &= -\frac{1}{\lambda_0}\ [1+\Theta]^{M+1} + K_0(0)\ [1+\gamma_0-\delta]^{M+1} \end{aligned}\right\} \quad t = M \qquad (4.7)$$

$$\left.\begin{aligned} X_0(t) &= \frac{1}{\lambda_0}\ (\Theta+\delta)\ (1+\Theta)^t \\ X(t) &= \frac{(\gamma_0-\Theta-\delta)}{\lambda_0}\ (1+\Theta)^t \end{aligned}\right\} \quad t = M+1, \ldots T-m-3 \qquad (4.8)$$

$$\left.\begin{aligned} X_0(t) &= 0 \\ X(t) &= \frac{1}{\lambda_0}(1+\Theta)^{T-m-2}(1-\delta)^{t-(T-m-2)} \end{aligned}\right\} t = T-m-2, \ldots T-2 \quad (4.9)$$

This solution is similar to that in Case 2a except that the labor constraint becomes binding at $t = (M+1)$. From period $(M+1)$ to period $(T-m-2)$ the entire labor force is employed in Sector I, thus reducing the rate of output of consumer goods to zero! Thus when T is sufficiently far off in the future, the optimal investment policy is to devote the entire output of machines to building up the stock of machines in Sector I during periods 0 to $(M-1)$; from period M to period $(T-m-3)$, gross investment is made in both sectors; from period $(T-m-2)$ to period $(T-2)$, the entire output of machines is invested in Sector II.

Maximizing consumption per worker at a selected future point in time thus leads to one of the following two consequences: Either consumption in the near future is emphasized at the expense of consumption in the far future, as in Case 1, or consumption at some far off point in time T is emphasized at the expense of consumption before T, as in Cases 2a and 2b. In either case, the maximized rate of consumption at T may not even be sustainable beyond T.

It might appear that one can modify the social objective considered above by maximizing the rate of consumption per worker at some T, subject to the constraints that the rate of consumption before T does not fall below a certain given magnitude and that the rate of consumption achieved at T be sustainable thereafter. But we have already shown in Section 2 that there is a unique maximum sustainable rate of consumption per worker. This is attained on the terminal path. This modified objective differs from the objective[1] in Section 3 in that a floor on the pre-terminal path rate of consumption has been imposed.

4.2. Derivations

We can formulate the problem of maximizing the rate of consumption per worker at time T (starting from given initial conditions and ignoring the rate of consumption prior to T) as the following linear programming problem, using the notation of Section 2.

Problem I

$$\text{Maximize } (1-\Theta)^{-T} \sum_{i=1}^{\infty} \gamma_i K_i'(T-1) \quad \text{subject to}$$

$$\sum_{i=1}^{\infty} \lambda_i K_i'(T-1) \leq (1+\Theta)^{T-1} \quad (4.10)$$

1. In Section 3 the objective was to minimize the time taken to reach the terminal path from given initial conditions.

$$-\sum_{t=0}^{T-2}(1-\delta)^{T-2-t}X_i(t)+K_i'(T-1)\leq K_i(0)(1-\delta)^{T-1}\quad i=1,2,\ldots \tag{4.11}$$

$$\sum_{i=0}^{\infty}X_i(t)-\gamma_0K_0'(t)\leq 0 \tag{4.12}$$

$$-\sum_{t'=0}^{t-1}(1-\delta)^{t-1-t'}X_0(t')+K_0'(t)\leq K_0(0)(1-\delta)^{t}\Big\}\ t=0,1,\ldots T-2 \tag{4.13}$$

$$\lambda_0K_0'(t)\leq(1+\theta)^t \tag{4.14}$$

$$K_0'(t),\ X_i(t),\ K_i'(T-1)\geq 0 \tag{4.15}$$

The maximand is the rate of consumption per worker at time T. Restraint (4.10) is the labor restraint at time T. (4.11) states that the stock of machines of any type in use at time (T-1) does not exceed that in existence. Restraints (4.12) – (4.14) refer to periods prior to (T-1). (4.12) states that the total gross investment in machines of various types at t does not exceed the rate of output of machines at t. (4.13) ensures that the stock of type 0 machines in use at t does not exceed that in existence. (4.14) is the labor restraint at t. It will be noticed that the variables $K_i'(t)$ i = 1, 2, ..., viz., the stock of machines in use of types other than 0, do not enter (4.14). The reason is that we have chosen to ignore the rate of consumption prior to T. Restraint (4.15) is the nonnegativity restraint on our variables.

We stated earlier that, along the optimal path, the gross addition to the stock of machines in Sector II will be divided among alternative types in constant proportions independent of t. In other words, along an optimal path, $\frac{X_i(t)}{X(t)}$ will be a constant independent of t. For showing this, we first have to reduce Problem I into an equivalent problem. We need to define a set of new variables. Let

$Y_i(T-1)$ = Addition (gross of depreciation of the initial stock) to the stock of machines of type i over the periods 0 - (T-2).

$I(T-1) = \sum_{i=1}^{\infty} Y_i(T-1)$. Clearly

$Y_i(T-1) = \sum_{t=0}^{T-2}(1-\delta)^{T-2-t}X_i(t)$. It will be recalled that, by

definition, $X(t) = \sum_{i=1}^{\infty} X_i(t)$. Now Problem I is equivalent to the following linear problem in a sense to be made more precise in the next paragraph.

Problem II

Maximize $(1+\Theta)^{-T} \sum_{i=1}^{\infty} \gamma_i K_i'(T-1)$ subject to

$$\sum_{i=1}^{\infty} \lambda_i K_i'(T-1) \leq (1+\Theta)^{T-1} \tag{4.16}$$

$$-Y_i(T-1) + K_i'(T-1) \leq K_i(0)(1-\delta)^{T-1} \quad i = 1, 2, \ldots \tag{4.17}$$

$$\sum_{i=1}^{\infty} Y_i(T-1) - I(T-1) = 0 \tag{4.18}$$

$$-\sum_{t=0}^{T-2} (1-\delta)^{T-2-t} X(t) + I(T-1) = 0 \tag{4.19}$$

$$X_0(t) + X(t) - \gamma_0 K_0'(t) \leq 0 \tag{4.20}$$

$$-\sum_{t'=0}^{t-1} (1-\delta)^{t-1-t'} X_0(t') + K_0'(t) \leq K_0(0)(1-\delta)^t \quad t = 0, 1, \ldots T-2 \tag{4.21}$$

$$\lambda_0 K_0'(t) \leq (1+\Theta)^t \tag{4.22}$$

$$K_0'(t),\ X_0(t),\ X(t),\ I(T-1),\ Y_i(T-1),\ K_i'(T-1) \geq 0 \tag{4.23}$$

It is easy to see that any feasible solution of Problem I, i.e., any set of values of the variables $X_i(t)$, $K_0'(t)$ and $K_i'(T-1)$ which satisfies (4.10) – (4.15), provides a feasible solution of Problem II. From any feasible solution of Problem II we can get a feasible solution of Problem I by setting[1]

$$X_i(t) = \left[\frac{Y_i(T-1)}{I(T-1)}\right] X(t) \quad \text{if} \quad I(T-1) > 0$$
$$= 0 \quad \text{if} \quad I(T-1) = 0$$

1. It is seen that our earlier assertion that $X(t)$ will be divided among alternative types in constant proportions independent of t is true.

The maximand is the same for both problems and the correspondence between the set of feasible solutions of the two problems leaves the values of the variables that enter the maximand unaltered. Hence Problems I and II are equivalent and it is sufficient to solve Problem II.

We now assert that Problem II can be decomposed into two linear programming problems. First, from (4.17) and (4.18) it is clear that increasing the value of I(T-1) can never decrease the optimal value of the maximand.[1] Second, the set of variables that enter the maximand, and the restraints (4.16) – (4.18) and the set of variables that enter the restraints (4.19) – (4.22), have only I(T-1) in common.[2] Hence we can decompose Problem II into the following two linear programming problems:

Problem IIa

$$\text{Maximize } I(T-1) = \sum_{t=0}^{T-2} (1-\delta)^{T-2-t} X(t) \quad \text{subject to}$$

$$X_0(t) + X(t) - \gamma_0 K_0'(t) \leq 0 \tag{4.24}$$

$$-\sum_{t'=0}^{t-1} (1-\delta)^{t-1-t'} X_0(t') + K_0'(t) \leq K_0(0)(1-\delta)^t \quad t = 0, 1, \ldots T-2 \tag{4.25}$$

$$\lambda_0 K_0'(t) \leq (1+\Theta)^t \tag{4.26}$$

$$X_0(t),\ X(t) \geq 0 \tag{4.27}$$

Problem IIb

$$\text{Maximize } (1-\epsilon)^{-T} \sum_{i=1}^{\infty} \gamma_i K_i'(T-1) \quad \text{subject to}$$

$$\sum_{i=1}^{\infty} \lambda_i K_i'(T-1) \leq (1+\Theta)^{T-1} \tag{4.28}$$

$$-Y_i(T-1) + K_i'(T-1) \leq K_i(0)(1-\delta)^{T-1} \tag{4.29}$$

$$\sum_{i=1}^{\infty} Y_i(T-1) - I_{max}(T-1) = 0 \tag{4.30}$$

$$K_i'(T-1),\ Y_i(T-1) \geq 0 \tag{4.31}$$

1. By assuming that there exists an i with λ_i sufficiently small, we can assert that the optimal value of the maximand is in fact increased by increasing I(T-1).

2. This assertion is possible only because we did not impose a floor on consumption at any $t < T$.

It is seen that Problem IIb involves the period (T-1) only, and its solution simply determines the distribution of the gross investment X(t) in Sector II between alternative types of machines. For, as we showed earlier, the $X_i(t)$ are to be determined by $X_i(t) = \left[\frac{Y_i(T-1)}{I_{max}(T-1)}\right] X(t)$. The time path of the variables X(t) and $X_0(t)$ can be obtained by solving Problem IIa. We present the solutions of Problem IIa only.

The optimal solutions to Problem IIa fall under three cases, as was mentioned earlier. We now present the solutions and establish their optimality by associating with each solution a set of nonnegative shadow prices and showing that under the chosen price system the solution is sustained. In the language of activity analysis, we need to show that under the chosen price system no activity makes a positive profit and all activities operated at a positive level break even. We use the following notation to translate these price conditions into algebraic terms:

q(t) : Shadow price of a unit of newly produced machine becoming available at time t.
r(t) : Rent per period of a unit of type 0 machine at t.
w(t) : Wages per period per unit of labor at t.

By $X_0(t)$, X(t), and $K_0'(t)$, let us denote respectively the activities (at t) of adding one machine to the stock of machines in Sector I, Sector II and the activity of producing machines. Let the same symbols represent the levels at which the corresponding activities are operated. Then the price conditions associated with an optimal solution can be written as follows:

$$\left.\begin{aligned} q(t) &\geq \sum_{t'=t+1}^{T-2} (1-\delta)^{t'-t} r(t') \quad \text{for all } t \\ &= \sum_{t'=t+1}^{T-2} (1-\delta)^{t'-t} r(t') \quad \text{if } X_0(t) > 0 \end{aligned}\right\} \quad (4.32)$$

$$\left.\begin{aligned} q(t) &\geq (1-\delta)^{T-1-t} \quad \text{for all } t \\ &= (1-\delta)^{T-1-t} \quad \text{if } X(t) > 0 \end{aligned}\right\} \quad (4.33)$$

$$\left.\begin{aligned} \gamma_0 q(t+1) &\leq \lambda_0 w(t) + r(t) \quad \text{for all } t \\ &= \lambda_0 w(t) + r(t) \quad \text{if } K_0'(t) > 0 \end{aligned}\right\} \quad (4.34)$$

The first part of (4.32) states that the activity $X_0(t)$ never makes a positive profit and the second part states that $X_0(t)$ breaks even whenever it is operated at a positive level. Another way of stating the economic meaning of (4.32) is to say that the cost of producing a new machine should be equal to the sum of the stream of returns (rents) from it in Sector I over the relevant period if it is to be added to the stock of machines in that sector. (4.33) and (4.34) express similar profit maximizing conditions with respect to activities $X(t)$ and $K_0'(t)$.[1] Now the optimal quantity solutions and the corresponding price solutions can be given.[2]

Case 1 : $T \leq (m+1)$. The solutions are

$$\left.\begin{aligned} X_0(t) &= 0 \\ X(t) &= \gamma_0 K_0(0)(1-\delta)^t \\ K_0'(t) &= K_0(0)(1-\delta)^t \end{aligned}\right\} \quad t = 0, 1, \ldots, T-2 \qquad (4.35)$$

$$q(t) = (1-\delta)^{T-1-t} \qquad t = 1, 2, \ldots, T-1 \qquad (4.36)$$

$$\left.\begin{aligned} r(t) &= \gamma_0[1-\delta]^{T-2-t} \\ w(t) &= 0 \end{aligned}\right\} \quad t = 0, 1, \ldots, T-2 \qquad (4.37), (4.38)$$

By direct substitution in (4.24) – (4.27), one can verify that the quantity solution is feasible. The price solution can be seen to sustain the quantity solution by substitution in (4.32) – (4.34). Therefore the above solution is optimal. As a check on our price computations, we can verify that the optimal value of the maximand is imputed to the scarce factors of production, i.e., the binding restraints. In other words, the following is seen to hold:

$$\sum_{t=0}^{T-2} (1-\delta)^{T-2-t} X(t) = \sum_{t=0}^{T-2} \left\{ K_0(0)(1-\delta)^t \right\} r(t)$$

$$= \gamma_0 (T-1) K_0(0)(1-\delta)^{T-2} .$$

---

1. The assumed production lag of one period is built into the price inequalities. In other words, outputs are evaluated at the price of the period in which they become available and inputs are evaluated at the prices of the period in which they are used.

2. The price of a machine in Sector II (irrespective of its type) available for use at time (T-1) is set equal to unity.

Case 2a : $T > (m+1)$ and $(T-m-2) \leq M$. The solutions are

$$\left.\begin{aligned} X_0(t) &= \gamma_0 K_0(0)\,(1+\gamma_0-\delta)^t \\ X(t) &= 0 \\ K_0'(t) &= K_0(0)\,(1+\gamma_0-\delta)^t \end{aligned}\right\} \quad t = 0, 1, \ldots, T-m-3$$

$$\left.\begin{aligned} X_0(t) &= 0 \\ X(t) &= \gamma_0 K_0(0)\,(1+\gamma_0-\delta)^{T-m-2}\,(1-\delta)^{t-(T-m-2)} \\ K_0'(t) &= K_0(0)\,(1+\gamma_0-\delta)^{T-m-2}\,(1-\delta)^{t-(T-m-2)} \end{aligned}\right\} \quad t = T-m-2, \ldots, T-2$$

$$\left.\begin{aligned} q(t) &= \gamma_0(m+1)\,(1-\delta)^m\,(1+\gamma_0-\delta)^{(T-m-1)-t} && t = 1, \ldots, T-m-2 \\ &= (1-\delta)^{T-1-t} && t = T-m-1, \ldots, T-1 \end{aligned}\right\}$$

$$\left.\begin{aligned} r(t) &= \gamma_0^2(m+1)\,(1-\delta)^m\,(1+\gamma_0-\delta)^{(T-m-2)-t} && t = 0, 1, \ldots, T-m-3 \\ &= \gamma_0(1-\delta)^{T-2-t} && t = T-m-2, \ldots, T-2 \end{aligned}\right\}$$

$$w(t) = 0 \qquad t = 0, 1, \ldots, T-2$$

Again it can be seen by substitution in the relevant inequalities that the above quantity solution is feasible and that it is sustained by the set of prices given above. Therefore the solution is optimal. One can verify also that the optimal value of the maximand is equal to the imputed value of the scarce factors. Stated in algebraic terms, we can show that

$$\sum_{t=0}^{T-2} (1-\delta)^{T-2-t}\,X(t) = \sum_{t=0}^{T-2} \left\{K_0(0)\,(1-\delta)^t\right\} r(t)$$

$$= \gamma_0(m+1)\,K_0(0)\,(1-\delta)^m\,(1+\gamma_0-\delta)^{T-m-2}$$

Case 2b : $T > (m+1)$ and $T-m-2 > M$. The solutions are

$$\left.\begin{aligned} X_0(t) &= \gamma_0\,K_0(0)\,(1+\gamma_0-\delta)^t \\ X(t) &= 0 \\ K_0'(t) &= K_0(0)\,(1+\gamma_0-\delta)^t \end{aligned}\right\} \quad t = 0, 1, \ldots, M-1$$

$$\left.\begin{aligned} X_0(t) &= \frac{1}{\lambda_0}\,[1+\Theta]^{M+1} - (1-\delta)\,K_0(0)\,(1+\gamma_0-\delta)^{M} \\ X(t) &= -\frac{1}{\lambda_0}\,(1+\Theta)^{M+1} + K_0(0)\,(1+\gamma_0-\delta)^{M+1} \\ K_0'(t) &= K_0(0)\,(1+\gamma_0-\delta)^{M} \end{aligned}\right\} t = M$$

$$\left.\begin{aligned} X_0(t) &= \frac{1}{\lambda_0}\,(\Theta+\delta)\,(1+\Theta)^{t} \\ X(t) &= \frac{(\gamma_0-\Theta-\delta)}{\lambda_0}\,(1+\Theta)^{t} \\ K_0'(t) &= \frac{(1+\Theta)^{t}}{\lambda_0} \end{aligned}\right\} t = M+1, \ldots, T-m-3$$

$$q(t) = (1-\delta)^{T-M-2}\,[1+\gamma_0-\delta]^{M+1-t} \quad t = 1, 2, \ldots, M$$

$$= (1-\delta)^{T-1-t} \quad t = M+1, \ldots, T-1$$

$$r(t) = \gamma_0(1-\delta)^{T-M-2}\,[1+\gamma_0-\delta]^{M-t} \quad t = 0, 1, \ldots, M$$

$$= 0 \quad t = M+1, \ldots, T-m-3$$

$$= \gamma_0(1-\delta)^{m}\,[(1-\delta) - \gamma_0 m] \quad t = T-m-2$$

$$= \gamma_0(1-\delta)^{T-m-2} \quad t = T-m-1, \ldots, T-2$$

$$w(t) = 0 \quad t = 0, 1, \ldots, M$$

$$= \frac{\gamma_0(1-\delta)^{T-t-2}}{\lambda_0} \quad t = M+1, \ldots, T-m-3$$

$$= \frac{1}{\lambda_0}\,(1-\delta)^{m}\,[\gamma_0(m+1) - (1-\delta)] \quad t = T-m-2$$

$$= 0 \quad t = T-m-1, \ldots, T-1$$

One can again verify by direct substitution in the relevant inequalities that the quantity solution given above is feasible and that it is sustained by the

above set of prices. As a computational check, one can verify that the optimal value of the maximand is imputed to the scarce resources, i.e., binding constraints. In other words, one can show that

$$\sum_{t=0}^{T-2} (1-\delta)^{T-2-t} X(t) = \sum_{t=0}^{T-2} \left\{ K_0(0)\,(1-\delta)^t \right\} r(t) + \sum_{t=0}^{T-2} (1-\Theta)^t\, w(t)$$

$$= K_0(0)\,(1-\delta)^{T-M-2}\,(1+\gamma_0-\delta)^{M+1}$$

$$+ \frac{(1+\Theta)^{T-m-2}\,(1-\delta)^m}{\lambda_0(\Theta+\delta)} \left[(\gamma_0-\Theta-\delta)(1-\delta)+\gamma_0(m+1)(\Theta+\delta)\right]$$

$$- \frac{\gamma_0(1+\Theta)^{M+1}(1-\delta)^{T-M-3}}{\lambda_0(\Theta+\delta)}$$

5. Social Welfare as a Discounted Sum of the Future Stream of Consumption.

5.1. Results

In this section we examine another social welfare function. Some authors view social welfare as an increasing function of the sum of the discounted stream of future consumption per worker.[1] Stated in the notation of Section 2, social welfare U(c) associated with a time path c is given by the following equation:

$$U(c) = \sum_{t=1}^{\infty} (1+\alpha)^{-t} c_t \qquad (5.1)$$

α is called the social rate of time preference. One important criticism against the use of U(c) for development planning is that U(c) implies a marginal rate of substitution between consumption at any two periods t_1 and t_2 which depends on the difference between t_1 and t_2 and is independent of the rates of consumption at t_1 and t_2. In this section, it is shown that this property of U(c) leads to the choice of time paths of consumption of extreme character.[2]

A society with U(c) as its welfare function discounts future consumption simply because it is available only in the future. Such behavior is often justified in the case of an individual because his life span is finite and he is subject to the process of aging, which might conceivably impair his capacity

1. See, for instance, Eckstein, op. cit., and Chenery, op. cit.

2. The results of this chapter are not new. Similar findings have been reported by Jan Tinbergen. Cf. Jan Tinbergen, "Optimum Savings and Utility Maximization Over Time," Econometrica 28 (April 1960), pp. 481-489.

for consumption. But this argument is not valid for a society or nation because it does not necessarily have a finite lifespan.

This is not to say that a society ought to treat a unit of consumption at two different points of time as perfect substitutes. We are arguing only against discounting the future simply because it is future. [1]

Since our purpose in this section is only to demonstrate the extreme character of the time paths that result from maximizing U(c) as given by (5.1), we drastically simplify the technology of production. Assume that no labor is needed either for producing machines or for producing consumer goods. Under this assumption, only one technique for producing consumer goods exists in Sector II, namely the technique which produces the maximum rate of consumer goods per machine. [2] We no longer need to distinguish between the stock of machines in existence and the stock in use since it will be inefficient to leave any machine idle. In the notation of Section 2, the problem can be stated as follows:

$$\text{Maximize } U(c) = \sum_{t=1}^{\infty} (1+\alpha)^{-t} c_t \quad \text{subject to}$$

$$(1+\Theta)^{t+1} c_{t+1} = \gamma_1 K_1(t) \tag{5.2}$$

$$K_i(t) = K_i(0)(1-\delta)^t + \sum_{t=0}^{t-1} (1-\delta)^{t-1-t'} X_i(t') \quad i = 0, 1 \tag{5.3}$$

$$X_0(t) + X_1(t) \leq \gamma_0 K_0(t) \tag{5.4}$$

$$X_0(t),\ X_1(t) \geq 0 \tag{5.5}$$

Equation (5.2) relates the rate of output of consumer goods during period t (becoming available for consumption in period (t+1)) to the stock of machines in Sector II. Equations (5.3) give the stock of machines of type 0 and 1 respectively. Inequality (5.4) states that the rate of gross investment in machines of type 0 and 1 together cannot exceed the rate of output of machines.

The solution to the above maximization depends on the value of the discount factor $(1+\alpha)$. When $(1+\beta) = (1+\alpha)(1+\Theta) \geq 1 + \gamma_0 - \delta$, the

---

1. Professor Maurice Dobb discusses the role of pure time preference in national planning quite extensively (Dobb, op. cit., Ch. 2).

2. Technique 1 yields the maximum rate of output of consumer goods per machine.

maximizing solution[1] is for $t \geq 0$

$$\left.\begin{aligned} X_0(t) &= 0 & X_1(t) &= \gamma_0 K_0(0)\,(1-\delta)^t \\ K_0(t) &= K_0(0)(1-\delta)^t & K_1(t) &= \left[(1-\delta)K_1(0) + \gamma_0 t K_0(0)\right](1-\delta)^{t-1} \\ c_{t+1} &= \gamma_1[1+\Theta]^{-t-1}\left[(1-\delta)\,K_1(0) + \gamma_0 t\,K_0(0)\right](1-\delta)^{t-1} \end{aligned}\right\} \quad (5.6)$$

This solution means that if the rate at which future consumption is discounted is greater than the maximum feasible rate of growth of the stock of machines per capita in the sector producing machines, the optimal policy is to invest the entire output of new machines in the sector producing consumer goods. In other words, the desire for current consumption is so great relative to future consumption that the society adopts a policy of permitting consumption per worker to decline steadily to zero from a certain point in time.

If $(1+\beta) < (1+\gamma_0-\delta)$, there is no maximizing solution because the maximand is unbounded. In other words, given any M, however large, we can find a time path c such that $U(c) > M$. Thus it is clear that, as a social objective, the attempts to maximize $U(c)$ leads to one of two extreme results: 1) either per capita consumption c_t steadily declines to zero beyond a certain time point, or 2) the process of maximization itself ceases to have any meaning with the maximand unbounded. It may appear that the removal of some of the simplifying assumptions on the technology of production modifies the extreme nature of this conclusion. But it is shown in Section 5.2 that the reintroduction of labor as a factor of production in the technology does not make any difference in the case $(1+\alpha)\ (1+\Theta) > 1 + \gamma_0 - \delta$. Consumption per worker along a time path which maximizes $U(c)$ still declines to zero. The only difference is that $U(c)$ is always bounded above as long as every technique of producing consumer goods uses nonzero amounts of labor services.

5.2. Derivations

Let us now solve the problem of maximizing $U(c) = \sum_{t=1}^{\infty} (1+\alpha)^{-t} c_t$ subject to

$$(1+\Theta)^{t+1} c_{t+1} = \gamma_1 K_1(t) \quad (5.7)$$

---

1. If $\beta = (\gamma_0-\delta)$, it is a matter of indifference how gross investment is divided between the two sectors. All feasible divisions yield the same value for $U(c)$. Therefore the maximization of $U(c)$ by itself does not provide any basis for choice between alternative feasible divisions.

$$K_i(t) = K_i(0)(1-\delta)^t + \sum_{t'=0}^{t-1} (1-\delta)^{t-1-t'} X_i(t') \quad i = 0, 1 \tag{5.8}$$

$$X_0(t) + X_1(t) \leq \gamma_0 K_0(t) \tag{5.9}$$

$$X_0(t),\ X_1(t) \geq 0 \tag{5.10}$$

Let us define a new parameter β by

$$(1+\alpha)(1+\Theta) = (1+\beta) \tag{5.11}$$

Using (5.7), (5.8), and (5.11) and omitting terms that are unessential to the choice of the $X_0(t)$ and $X_1(t)$ that maximize U(c), we can rewrite the problem as:

$$\text{Maximize} \sum_{t=0}^{\infty} (1+\beta)^{-t} X_1(t) \quad \text{subject to}$$

$$-\gamma_0 \sum_{t'=0}^{t-0} (1-\delta)^{t-1-t'} X_0(t') + X_0(t) + X_1(t) \leq \gamma_0 K_o(0)(1-\delta)^t \tag{5.12}$$

$$X_0(t),\ X_1(t) \geq 0 \tag{5.13}$$

The problem is in a standard linear programming form, and the optimality of a feasible solution to a linear programming problem can be tested by associating a set of shadow prices with the restraints and checking whether the given feasible solution is sustained by the price system.[1] Let u(t) be the shadow price associated with restraint (5.12). Then a feasible solution is optimal if the associated values of u(t) satisfy:

$$\left.\begin{aligned} u(t) &\leq \gamma_0 \sum_{t'=t+1}^{\alpha} (1-\delta)^{t'-t-1} u(t') \qquad \text{for all } t \geq 0 \\ &= \gamma_0 \sum_{t'=t+1}^{\infty} (1-\delta)^{t'-t-1} u(t') \qquad \text{if } X_0(t) > 0 \end{aligned}\right\} \tag{5.14}$$

1. Standard linear programming problems involve a finite number of variables. We have here a denumerably infinite number of variables $X_0(t)$, $X(t)$. It is shown in the Appendix that our procedure of testing optimality by prices works in this case as well.

$$\left.\begin{aligned} u(t) &\leq (1+\beta)^{-t} && \text{for all } t \geq 0 \\ &= (1+\beta)^{-t} && \text{if } X_1(t) > 0 \end{aligned}\right\} \quad (5.15)$$

We mentioned in Section 5.1 that a maximizing solution exists only when $\beta \geq (\gamma_0 - \delta)$. The optimal solution and the associated prices in this case are:

$$\left.\begin{aligned} X_0(t) &= 0 && u(t) = (1+\beta)^{-t} \\ X_1(t) &= \gamma_0 K_0(0)\,(1-\delta)^t \end{aligned}\right\} \quad (5.16)$$

By substituting these values in (5.12) – (5.15), one can verify that the solutions (5.16) are optimal.[1] Consumption per worker at time t along the optimal path is given by $c_t = (1+\Theta)^{-t}\ \gamma_1 K_1(t-1)$. Substituting for $K_1(t-1)$ from (5.8) and using (5.16) we have

$$c_t = (1+\Theta)^{-t}\ (1-\delta)^{t-2}\,[(1-\delta)K_1(0) + \gamma_0(t-1)\,K_0(0)]$$

It is seen that c_t steadily declines to zero from a certain point in time.

In order to show that the maximand is unbounded when $\beta < (\gamma_0 - \delta)$, consider the following class of feasible solutions:

$$\left.\begin{aligned} X_0(t) &= \gamma_0\, K_0(0)\,[1+\gamma_0-\delta]^t \\ X_1(t) &= 0 \end{aligned}\right\} \quad t = 0, 1, 2, \ldots, T-1 \quad (5.17)$$

$$\left.\begin{aligned} X_0(t) &= 0 \\ X_1(t) &= \gamma_0\, K_0(0)\,[1+\gamma_0-\delta]^T\ (1-\delta)^{t-T} \end{aligned}\right\} \quad t = T, T+1, \ldots \quad (5.18)$$

It is easy to verify that for each $T \geq 1$, (5.17) and (5.18) provide a feasible solution. The value of the maximand for such a solution is given by

$$\sum_{t=0}^{\infty} (1+\beta)^{-t} X_1(t) = \gamma_0\, K_0(0)\ (1+\gamma_0-\delta)^T \sum_{t=T}^{\infty} (1+\beta)^{-t}\ (1-\delta)^{t-T}$$

$$= \frac{\gamma_0 K_0(0)}{(\beta+\delta)} \left[\frac{1+\gamma_0-\delta}{1+\beta}\right]^T \quad (5.19)$$

---

1. In the case $(\gamma_0 - \delta) = \beta$ it is immaterial how investment is divided between $X_0(t)$ and $X_1(t)$. See footnote on page 101.

Given that $\beta < (\gamma_0 - \delta)$, it is clear that $\left[\frac{1+\gamma_0-\delta}{1+\beta}\right]^T$ increases indefinitely as T increases. Therefore the maximand can be made to exceed any value M, however large, by choosing a feasible path given by (5.17) and (5.18) for some sufficiently large value of T.

Thus far we have been assuming that no labor was needed to produce either machines or consumer goods. Let us now relax this assumption, thereby reintroducing the whole range of techniques of production of consumer goods. The problem can now be stated as:

$$\text{Maximize } U(c) = \sum_{t=1}^{\infty} (1+\alpha)^{-t} c_t \quad \text{subject to (for } t \geq 0)$$

$$(1+\Theta)^{t+1} c_{t+1} = \sum_{i=1}^{\infty} \gamma_i K_i'(t) \tag{5.20}$$

$$K_i(t) = K_i(0)(1-\delta)^t + \sum_{t'=0}^{t-1} (1-\delta)^{t-1-t'} X_i(t') \quad i = 0, 1, 2, \ldots, \tag{5.21}$$

$$K_i'(t) - K_i(t) \leq 0 \tag{5.22}$$

$$\sum_{0}^{\infty} \lambda_i K_i'(t) \leq (1+\Theta)^t \tag{5.23}$$

$$\sum_{0}^{\infty} X_i(t) - \gamma_0 K_0'(t) \leq 0 \tag{5.24}$$

$$X_i(t),\ K_i'(t),\ \geq 0 \qquad i = 0, 1, \ldots \tag{5.25}$$

The maximand is, as before, the sum of the discounted stream of future consumption. Equation (5.20) states that the rate of output of consumer goods becoming available at (t+1) is the sum of the output produced with various techniques. Equations (5.21) give the stocks of various types of machines in existence.

The exact form of the solution to the above problems depends on the initial conditions, the discount factor, and the parameters of the technology. We present here only two cases. Let us first define a time function g(t) as follows:

$$g(t) = \left[(1-\delta)\left(\sum_{i=0}^{\infty} \lambda_i K_i(0)\right) + \left(\lambda_1 \gamma_0 K_0(0)t\right)\right](1-\delta)^{t-1} \tag{5.26}$$

It can be seen that g(t) represents the potential level of employment in period t if (a) the existing stocks of each type of machine are used, and (b), in each period t, the entire output of new machines is added to the stock of type 1 machines. The behavior of g(t) is relevant for distinguishing the two cases to be discussed.

As in earlier sections, we prove the optimality of a feasible solution by associating with it a set of nonnegative prices which sustains it. In the notation of Section 2.2, we have to show that[1]

$$\left.\begin{aligned} \gamma_0\, q(t+1) &\leq \lambda_0\, w(t) + r(t) \qquad \text{for all } t \\ &= \lambda_0\, w(t) + r(t) \qquad \text{if } K_0'(t) > 0 \end{aligned}\right\} \tag{5.27}$$

$$\left.\begin{aligned} \gamma_i\, p(t+1) &\leq \lambda_i w(t) + r_i(t) \quad \text{for all } t \text{ and } i \neq 0 \\ &= \lambda_i w(t) + r_i(t) \quad \text{if } K_i'(t) > 0 \end{aligned}\right\} \tag{5.28}$$

$$\left.\begin{aligned} q(t+1) &\geq \sum_{t'=t+1}^{\infty} (1-\delta)^{t'-t-1}\, r_i(t') \quad \text{for all } t \\ &= \sum_{t'=t+1}^{\infty} (1-\delta)^{t'-t-1}\, r_i(t') \quad \text{if } X_i(t) > 0 \end{aligned}\right\} \tag{5.29}$$

Case 1 : $g(t) \leq (1+\Theta)^t$ for all t and $\beta \geq (\gamma_0 - \delta)$. The quantity and price solutions are:

$$\left.\begin{aligned} X_1(t) &= \gamma_0 K_0(0)\,(1-\delta)^t \\ K_1(t) &= K_1'(t) = \left[(1-\delta)\, K_1(0) + t\,\gamma_0 K_0(0)\right](1-\delta)^{t-1} \end{aligned}\right\} \text{for all } t \tag{5.30}$$

$$\left.\begin{aligned} X_i(t) &= 0 \\ K_i'(t) &= K_i(t) = K_i(0)\,(1-\delta)^t \end{aligned}\right\} \text{for } i \neq 1 \text{ and for all } t \tag{5.31}$$

---

1. It can be seen that all prices are present values and that the present value of a unit of consumer good available at t is determined by the maximand as $[1+\beta]^{-t}$ where $(1+\beta) = (1+\alpha)(1+\Theta)$.

$$\left.\begin{aligned} p(t) &= [1+\beta]^{-t} \\ w(t) &= 0 \\ r_0(t) &= \frac{\gamma_0\gamma_1[1+\beta]^{-t-1}}{(\beta+\delta)} \\ r_i(t) &= \gamma_i[1+\beta]^{-t-1} \\ q(t) &= \frac{\gamma_1[1+\beta]^{-t}}{(\beta+\delta)} \end{aligned}\right\} \text{ for all } t \geq 0 \qquad (5.32)$$

Given that $g(t) \leq (1+\Theta)^t$, it can be verified that the quantity solutions (5.30) and (5.31) satisfy (5.21) – (5.25). The condition that $\beta \geq (\gamma_0-\delta)$ ensures that the prices (5.32) satisfy (5.27) – (5.29). Therefore the feasible solution given by (5.21) – (5.25) is optimal. The rate of consumption per worker at t as per this solution is

$$c_t = (1+\Theta)^{-t}(1-\delta)^{t-2}\left[(1-\delta)\left(\sum_{i=1}^{\infty}\gamma_i K_i(0)\right) + \gamma_1\gamma_0(t-1)K_0(0)\right]$$

It is obvious that $c_t \to 0$ as $t \to \infty$.

Case 2 :

(a) There exists a t_1 such that

$$g(t_1) \leq (1+\Theta)^{t_1} \text{ and } g(t_1+1) > (1+\Theta)^{t_1+1}$$

(b) $$g(t_1+1) - \gamma_0(\lambda_1-\lambda_2)\,K_0(0)\,(1-\delta)^{t_1} \leq (1+\Theta)^{t_1+1}$$

(c) $$\frac{\gamma_0(\lambda_1\gamma_2-\lambda_2\gamma_1)}{\lambda_0(\gamma_1-\gamma_2)} > (\beta+\delta) > \frac{\lambda_1\gamma_0}{\lambda_0} > (1-\delta)$$

Notice that (c) implies that $(\beta+\delta) \geq \gamma_0$ since $\lambda_1 \geq \lambda_0$ by assumption. The quantity solutions in this case are: [1]

1. The prices that sustain this quantity solution can be derived quite easily.

$$K_i'(t) = K_i(0)(1-\delta)^t + \sum_{t'=0}^{t-1} (1-\delta)^{t-1-t'} X_i(t') \quad \text{for all } i \text{ and for all } t \tag{5.33}$$

$$X_i(t) = 0 \quad \text{for all } i \neq 1, 2 \text{ and for all } t \tag{5.34}$$

$$\left.\begin{aligned} X_1(t) &= \gamma_0 K_0(0)(1-\delta)^t \\ X_2(t) &= 0 \end{aligned}\right\} t = 0, 1, 2, \ldots, t_1 - 1 \tag{5.35}$$

$$\left.\begin{aligned} X_1(t_1) &= \frac{(1-\Theta)^{t_1+1} + \gamma_0(\lambda_1 - \lambda_2) K_0(0)(1-\delta)^{t_1} - g(t_1+1)}{(\lambda_1 - \lambda_2)} \\ X_2(t_1) &= \frac{g(t_1+1) - (1+\Theta)^{t_1+1}}{(\lambda_1 - \lambda_2)} \end{aligned}\right\} \tag{5.36}$$

$$\left.\begin{aligned} X_1(t) &= \frac{(\Theta+\delta)(1+\Theta)^t - \lambda_2 \gamma_0 K_0(0)(1-\delta)^t}{(\lambda_1 - \lambda_2)} \\ X_2(t) &= \frac{\lambda_1 K_0(0)(1-\delta)^t - (\Theta+\delta)(1+\Theta)^t}{(\lambda_1 - \lambda_2)} \end{aligned}\right\} t = t_1, \ldots t_2 \tag{5.37}$$

where t_2 is defined by

$$\left.\begin{aligned} \lambda_1 K_0(0)(1-\delta)^{t_2} &\geqq (\Theta+\delta)(1+\Theta)^{t_2} \\ \lambda_1 K_1(0)(1-\delta)^{t_2+1} &< (\Theta+\delta)(1+\Theta)^{t_2+1} \end{aligned}\right\} \tag{5.38}$$

$$\left.\begin{aligned} X_1(t) &= \gamma_0 K_0(0)(1-\delta)^t \\ X_2(t) &= 0 \end{aligned}\right\} t \geqq t_2 + 1 \tag{5.39}$$

It is easy to verify that this solution is feasible, given that (a) and (b) hold. Given (c), it can be shown that it is optimal as well. In this case also, it can be seen that consumption per worker goes to zero as $t \to \infty$. For

$$c_{t+1} = (1-\delta)\left\{\sum_{1}^{\infty} \gamma_i K_i'(t_2)\right\} + \gamma_1\gamma_0 K_0'(t_2)\,(t-t_2)\,(1-\delta)^{t-t_2-1} \quad t \geq t_2 \qquad (5.40)$$

where $K_i'(t_2)$ are given by (5.33).

It is obvious from (5.40) that $\lim_{t \to \infty} c_{t+1} = 0$.

6. Some Characteristics of Efficient Time Paths

6.1. Results

In this section some common characteristics of all efficient time paths are derived. Previously,[1] an efficient time path c such that there is no other feasible time path c' which satisfies

$$\left.\begin{aligned} c_t' &\geq c_t \quad \text{for all } t \\ c_t' &> c_t \quad \text{for at least one } t \end{aligned}\right\} \qquad (6.1)$$

The following theorem states a necessary and sufficient condition for a feasible path to be efficient.

Theorem 1

A feasible time path c is efficient if and only if, for all h, there is no feasible time path c' which satisfies

$$\left.\begin{aligned} c_t' &\geq c_t \quad \text{for all } t \leq h \\ c_t' &> c_t \quad \text{for some } t \leq h \\ c_t' &= c_t \quad \text{for all } t > h \end{aligned}\right\} \qquad (6.2)$$

The meaning of this theorem is obvious: A feasible path c is efficient if and only if there is no feasible time path c' which provides at least as much consumption per worker as c at any t within a finite time horizon h, more consumption per worker than c at some $t' \leq h$, and exactly as much consumption per worker as c at any t beyond h.

The following result relates the utilization of the existing stocks of various types of machines with efficiency.[2]

---

1. See Definition 3, Section 2.

2. It is worth recalling that machines of type i ($i \geq 1$) are used to produce consumer goods with technique i. The techniques are numbered in such a way that the rate of output of consumer goods per unit of labor increases with i.

Theorem 2

Along any efficient path c, if at any t the existing stock of type j ($j > 1$) machines is not fully utilized, then the stock in use of type i machine will be zero for $1 \leq i < j$.

It is obvious why Theorem 2 should be true, for labor is more productive with type j machines than with type i for $j > i \geq 1$. If along any feasible path c at any t (a) the existing stock of type j machines is not fully utilized, and (b) the stock in use of type i machine is positive for some i in $1 \leq i < j$, then, by transferring labor from employment with type i machines to employment with type j machines, one can increase the rate of output of consumer goods at time t without reducing that at any other $t' \neq t$. Therefore c cannot be efficient.

Corollary 2.1

Along any efficient path c, the following can be true in any given period for, at the most, one type of machine: The stock in use is greater than zero but less than the stock in existence.

Corollary 2.2

Along any efficient path c, at least one of the following must hold at any t

(a) The existing stock of all types of machines which produce consumer goods are fully utilized.

(b) The available labor supply is fully utilized.

This corollary expresses the simple idea that it is inefficient to keep both men and machines idle.

The following theorem provides a necessary condition for positive gross investment to be made in a type of machine which produces consumer goods.

Theorem 3

On an efficient path c with $c_t > 0$ for all t, positive gross investment in type i machines ($i \geq 1$) will be made at any t only if the existing stock of machines of type i will be fully utilized at at least one $t' > t$.

The economic meaning of this theorem is trivial. It is inefficient to add to the existing stock of a type of machine which is never going to be fully utilized.

6.2. Derivations

Let us now prove these results.[1]

Theorem 1

A feasible time path c is efficient if and only if, for all h, there

1. Theorem 1 corresponds to Lemma 1 of Malinvaud's article (Malinvaud, op. cit., p. 244).

is no other feasible time path c' which satisfies

$$\left.\begin{aligned} c'_t &\geqq c_t \quad \text{for all} \quad t \leqq h \\ c'_t &> c_t \quad \text{for some} \quad t < h \\ c'_t &= c_t \quad \text{for all} \quad t > h \end{aligned}\right\} \tag{6.3}$$

The necessity part of the theorem is a direct consequence of the definition of efficiency. To prove the sufficiency part we first state and prove a lemma.

Lemma

If a time path c is feasible, then any time path c' such that $c'_t \leqq c_t$ for all t is also feasible.

Given that c is feasible, it is clear that c'_t can be obtained from c by suitably reducing the stock in use of machines of various types. Therefore c' is also feasible.

We can now prove the sufficiency part of Theorem 1. Suppose that, contrary to the assertion of the theorem, c is not efficient. Then there exists a feasible c' such that

$$c'_t \geqq c_t \quad \text{for all } t \quad \text{and}$$
$$c'_t > c_t \quad \text{for at least one } t \text{ say } (t = h).$$

Now define a time path c'' as follows

$$\left.\begin{aligned} c''_t &= c'_t \quad t \leqq h \\ c''_t &= c_t \quad t > h \end{aligned}\right\} \tag{6.4}$$

Clearly $c''_t \leqq c'_t$ for all t. Therefore, given that c' is feasible, c'' is also feasible. Now c'' violates the hypothesis of the theorem, thus bringing about a contradiction. Therefore c is efficient.

Theorem 2 and its corollaries and Theorem 3 are easily proved[1] by using the properties of the price system that can be associated with an efficient path. The existence of a sustaining price system associated with any efficient path has been proved by Malinvaud.[2] The existence theorem can

---

1. The same results can be also proved more directly without the use of efficiency prices. But the direct proofs are lengthy.

2. Malinvaud, op. cit., Theorem 1, p. 245. This theorem is the converse of the Lemma we used in proving the efficiency of a given feasible path in Section 2.2. It must be remembered that we are extending Malinvaud's results to the case where there is an infinite number of commodities in each period.

be stated as follows, using the notation of Section 2.

Given an efficient path c, there exists a set of nonnegative prices $p(t)$, $q(t)$, $r_i(t)$, and $w(t)$ $(t \geq 0)$ such that:

(a) $\gamma_i p(t+1) \leq \lambda_i w(t) + r_i(t)$ for all $i \geq 1$

$= \lambda_i w(t) + r_i(t)$ if $K_i'(t) > 0$

(b) $\gamma_0 q(t+1) \leq \gamma_0 w(t) + r_0(t)$

$= \lambda_0 w(t) + r_0(t)$ if $K_0'(t) > 0$

(c) $r_i(t) = 0$ if for any $i \geq 0$ $\quad 0 < K_i'(t) < K_i(t)$

(d) $w(t) = 0$ if the available supply of labor is not fully utilized at any t

(e) $q(t+1) \geq \sum_{t'=t+1}^{\infty} (1-\delta)^{t'-t-1} r_i(t')$ for all $i \geq 0$

$= \sum_{t'=t+1}^{\infty} (1-\delta)^{t'-t-1} r_i(t')$ if $X_i(t) > 0$

(f) The cost of inputs, evaluated in terms of the above prices for each period t, is minimum for the input combination associated with c, among all input combinations for that period which permit the same consumption per worker as in c for periods $t' \geq t$.

It is easily seen that properties (a) and (b) of the set of prices imply that the quantity solutions $K_i'(t)$ of c are profit maximizing levels, given the set of prices, of the activities that produce consumer goods and machines. Properties (c) and (d) state that unused resources earn zero rentals. Property (e) covers profit maximization in investment activities. Property (f) is needed only when we wish to prove the efficiency of a given feasible solution through prices.[1] It is possible that in certain cases one can associate with an inefficient path a set of prices with properties (a) - (e).

Theorem 2

On an efficient path c, if $0 < K_j'(t) < K_j(t)$ for some $j > 1$ and

1. For a discussion of the property (f) see Tjalling C. Koopmans, Three Essays on the State of Economic Science (1st ed.; New York: McGraw-Hill Book Company, Inc.), pp. 110-112.

and some $t \geq 0$, then $K_i'(t) = 0$ for all $i \leq j$.

Suppose, on the contrary, $K_i'(t) > \bar{0}$ for some $i < j$, then by (a) we have

$$\begin{aligned} \lambda_i w(t) + r_i(t) &= \gamma_i p(t+1) \\ \lambda_j w(t) + r_j(t) &= \gamma_j p(t+1) \end{aligned} \tag{6.5}$$

By (c) we know that $r_j(t) = 0$. Now $\frac{\lambda_i}{\gamma_i} > \frac{\lambda_i}{\gamma_j}$ and $p(t+1) > 0$ for all t since the consumer good is the only desired commodity in our model and saturation in consumption is ruled out. If (6.5) were to hold, then $r_i(t) < 0$, violating the nonnegative character of the efficiency prices. Thus $K_i'(t) = 0$ for all i in $1 \leq i < j$, given that $0 < K_j'(t) < K_j(t)$. It is also clear that $0 < K_j'(t) < K_j(t)$ can hold for, at the most, one $j \geq 1$, proving Corollary (2.1).

Corollary 2.2

Along any efficient path c, at least one of the following holds: for each t :

$$\sum_{i=0}^{\infty} \lambda_i K_i'(t) = (1+\theta)^t \tag{6.6}$$

$$K_i'(t) = K_i(t) \quad \text{for all } i \geq 1 \tag{6.7}$$

If strict inequalities (<) held in (6.6) and (6.7), (a) and (c) would imply $w(t) = 0$ and $r_i(t) = 0$ for all $i \geq 1$. These, in turn, would imply $\lambda_i w(t) + r_i(t) < \gamma_i p(t+1)$, thus violating a part of (a).

Theorem 3

On an efficient path c with $c_t > 0$ for all t, $X_i(t_1) > 0$ for some $i \geq 1$ and some t_1 only if $K_i'(t) = K_i(t)$ for some $t' > t_1$.

Suppose, on the contrary, $K_i'(t) < K_i(t)$ for all $t > t_1$; then, by (c), $r_i(t) = 0$ for all $t > t_1$. Now given that $X_i(t_1) > 0$, by (e) we know that

$$q(t_1+1) = \sum_{t'=t_1+1}^{\infty} (1-\delta)^{t'-t-1} r_i(t') = 0 . \tag{6.8}$$

Another implication of (6.8) is that $r_j(t') = 0$ for all j and for all $t' > t_1$. For if $r_j(t'') > 0$ for some $t'' < t_1$, then for such a j

$$q(t_1+1) < \sum_{t'=t+1}^{\infty} (1-\delta)^{t'-t-1} r_j(t')$$

This violates (e). Therefore $r_j(t') = 0$ for all j and for all $t' > t_1$.

Now given $c_t > 0$ for all t, it is clear that for each t $K_j^!(t) > 0$ for some $j \geq 1$. This implies that for such a j,

$$\lambda_j w(t) + r_j(t) = \gamma_j p(t+1) \tag{6.10}$$

Now consider a $t' > t_1$. We have already shown that $r_j(t') = 0$ for all j. Hence (6.10) implies

$$\lambda_j w(t') = \gamma_j p(t'+1) \tag{6.11}$$

Consider an $i > j$. For any i, $r_i(t') = 0$. Now (6.11) implies that

$$\lambda_i w(t') < \gamma_i p(t'+1) \quad \text{since} \quad \frac{\lambda_i}{\gamma_i} < \frac{\lambda_j}{\gamma_j}$$

Therefore $\lambda_i w(t') + r_i(t') < \gamma_i p(t'+1)$, contradicting (a), thus proving the theorem.

7. Summary and Qualifications

In this Section, the conclusions of the previous sections are summarized and qualified. In Section 1 we reviewed some important contributions to the problem of choice of investment criteria.

In Section 2 we considered a simple dynamic model of an economy with two sectors, Sector I producing machines and Sector II producing a homogeneous consumer good. The former used a single fixed combination of machines and labor to produce various types of machines. In the latter sector, many alternative techniques of production of differing labor intensities were available. It was demonstrated that there exists a maximum sustainable rate of consumption per worker for such an economy. This rate is achieved along a "terminal path." The terminal path was shown to be a balanced growth path in which the stocks of machines of various types grew at the same geometric rate as the labor force. While on the terminal path, only one of the many alternative techniques was used in Sector II at every point of time, and consumption per worker was constant at the maximum sustainable level.

In Section 3, the time path was derived which minimizes the time taken to reach the terminal path and remains on it thereafter, starting from given initial stocks of various types of machines.[1] This derived path implied a continuous decline in consumption per worker for a finite interval of time.

In Sections 4 and 5 we examined time paths that result from maximizing two specific social welfare functions: (a) consumption per worker at some chosen point of time T in the future and (b) the discounted sum of the future stream of consumption per worker. We concluded that these

1. We considered only the approaches to the terminal path from below.

time paths led to two possible results: (1) consumption per worker declines after a certain point of time, or (2) consumption at some point in time is maximized at the expense of consumption before that point and without any sustainability requirement on consumption beyond that point.

In Section 6 some common characteristics of efficient time paths were established.

It should be clear from our discussion in Sections 3 – 5 that it is inherently difficult to arrive at an analytically convenient social welfare function, the maximization of which yields a satisfactory time path of consumption per worker. A modified version of the objective considered in Section 3, such as minimizing the time taken to reach the terminal path subject to the constraint that consumption per worker never falls below a certain floor at any point in time, is worth exploring. The most serious analytical difficulty in this approach is the problem of the choice of techniques of production of consumer goods at each point in time prior to the attainment of the terminal path. This problem disappears if the constraint of a floor in consumption is removed.

Even if a satisfactory solution could be found for the choice of a social welfare function, our model has to be generalized in a number of directions before it can be used for a discussion of a realistic planning problem. We considered a closed economy which produced the machines it needed. The underdeveloped areas of today have no heavy industry to speak of and have to rely on imports from advanced countries for industrialization. More importantly, the introduction of international trade could conceivably alter the choice of techniques of production. An additional constraint in the form of the availability of foreign exchange will have to be introduced.

We made the rather convenient but unrealistic assumption that labor was homogeneous and can be transferred from one technique to another without incurring any cost. It is well known that one of the most serious obstacles to economic growth in underdeveloped areas is the lack of skilled manpower. A realistic model will have to include training activities, the input of which is unskilled labor and with trained workers being the output.

Investment in training and education is one form of investment in social overhead capital. Other examples are investment in transportation networks, power supply, etc. Investment in social overhead capital may yield increasing returns to scale. This analytically complex phenomenon has to be recognized in a development model.

The assumption that the labor force grows at a constant geometric rate per period needs revision. The rate of growth of population may depend upon the level and rate of growth of consumption, and as such it may be one of the choice variables.

We assumed a simple and unchanging technology for our economy. It is clear that in a model which covers an infinite time horizon, unchanging technology is a restrictive assumption. A thorough understanding of the process by which technical change is brought about is essential before

we can include this aspect in our model.

We conclude by expressing a hope that, in spite of the above-mentioned qualifications, this study has contributed towards a clearer understanding of the problem of the choice of investment criteria.

Appendix

Consider the following linear programming problem involving a finite number of variables and restraints:

$$\text{Minimize} \quad cx \qquad (A.1)$$

$$\text{Subject to} \quad Ax \geqq b \qquad (A.2)$$

$$x \geqq 0 \qquad (A.3)$$

where c is a row vector $(c_1, \ldots, c_n)$, x a column vector $(x_1, \ldots, x_n)'$, A is a mxn matrix and b a column vector $(b_1, b_2, \ldots, b_m)'$. Then the following theorem holds:

Theorem $\overset{\circ}{x}$ is optimal if we can find a row vector $\overset{\circ}{u} = (\overset{\circ}{u}_1, \ldots, \overset{\circ}{u}_m)$ such that

$$\overset{\circ}{x} \geqq 0 \qquad (A.4)$$

$$\overset{\circ}{u} \geqq 0 \qquad (A.5)$$

$$A\overset{\circ}{x} \geqq b \qquad (A.6)$$

$$\overset{\circ}{u} A \leqq c \qquad (A.7)$$

$$c\,\overset{\circ}{x} \leqq \overset{\circ}{u}\, b \qquad (A.8)$$

Proof: Consider any feasible x, i.e., any x which satisfies (A.2) and (A.3).

Since $\overset{\circ}{u} \geqq 0$ and $Ax \geqq b$ it is clear that $\overset{\circ}{u}\, Ax \geqq \overset{\circ}{u}\, b$

Since $\overset{\circ}{u}\, A \leqq c$ and $x \geqq 0$ we have $\overset{\circ}{u}\, Ax \leqq c\, x$

Therefore $c\, x \geqq \overset{\circ}{u}\, b \geqq c\, \overset{\circ}{x}$ showing that $\overset{\circ}{x}$ is optimal.

Suppose, now, that we admit a countably infinite number of variables, keeping the number of restraints finite. It is clearly seen that if each of the infinite sums in Ax, $\overset{\circ}{u}\, A$ and c x is convergent for all feasible x, the above proof that (A.4) – (A.8) are sufficient for the optimality of $\overset{\circ}{x}$ holds.

Summary: This essay presents a model of cash and security holdings of businesses. The model is tested on individual firm data, namely, a sample of large U.S. manufacturing corporations during the period 1947–1956. Firms are assumed to only partially adjust their holdings of cash and securities to desired levels within any year. Determinants of desired cash and security holdings are the bill rate, the level of transactions, and other current accounts of firms. It was found that the lagged adjustment model provided a reasonable description of behavior, and that the signs on all explanatory variables, with the exception of the bill rate on securities, were as expected. As the models employ separate intercepts for each firm, the emphasis of the analysis is on short-run adjustments. It was found that firms substitute freely between the use of securities and short-term bank loans to finance periodic increases in inventories and net receivables. It was also found that the accounts receivable of large corporations are more responsive to the bill rate than the accounts payable of large firms, so that large firms lend substantially in response to interest rate increases. Finally, securities, rather than cash, appear to bear the brunt of the short-run adjustments of firms to changes in sales and inventories.

This essay was submitted to Yale University in 1961 in partial fulfillment of the requirements for the degree of Doctor of Philosophy, to be awarded in 1962. The author is currently Assistant Professor of Economics at Yale University and is on the staff of the Cowles Foundation for Research in Economics at Yale.

Dissertation Committee: Professors Arthur M. Okun (Chairman), Raymond P. Powell, and James Tobin.

*

I would like to thank Edward Budd, William Brainard, Donald Hester, Richard Porter, and T. N. Srinivasan of Yale University, who provided many useful comments at various points in the study. I would especially like to thank Arthur Okun and James Tobin, who were members of my dissertation committee and who provided help through all stages of the research. The Ford Foundation and the Cowles Foundation for Research in Economics were kind enough to finance my research and computing expenses. Mr. Daniel Brill and Miss Eleanor Pratt of the Board of Governors of the Federal Reserve System were very cooperative about my use of the sample data collected by the Board of Governors. Finally, I would like to thank my wife, Wilma, whose contributions were many.

A.W.H.

AN EMPIRICAL STUDY OF CASH, SECURITIES, AND OTHER CURRENT ACCOUNTS OF LARGE CORPORATIONS

Alan W. Heston

1. Scope and Method of the Study

Non-financial corporations hold about one-fourth of the currency and demand deposits in the United States and about one-sixth of the total federal government marketable securities. In 1955 this sector created for itself roughly 50 billion dollars in credit as accounts payable. In addition net receivables, i.e., receivables minus payables, of this sector provided another 30 billion dollars in credit to non-corporate businesses, governments and consumers. An understanding of the short-run fluctuations of these and the other current accounts of corporations is important to an understanding of fluctuations in the economy as a whole.

This study concentrates on explaining short-period movements in cash and security holdings of an important segment of the corporate sector, large non-financial corporations. We will also be concerned with variations in the other current accounts of firms, and in particular with short-term bank loans and net receivables. Though some of our results relate to long-run behavior of large corporations, the study is orientated toward explaining short-run fluctuations in these accounts.

Any explanation of cash and security holdings will probably involve the level of transactions and the interest rate. In Section 2 a selected group of theoretical and empirical studies concerning the relation of cash balances, transactions, and the interest rate are reviewed. Section 2 also presents a theoretical discussion of the variables entering into the models of cash and securities. Section 3 presents empirical tests of propositions about the relationship of cash and securities and the other current accounts. Empirical tests of more detailed models of cash and securities are presented in Section 4. In Section 5, the empirical results are analysed and their implications for monetary policy are discussed.

This study is empirically oriented. Because the models are a liberal dose of experimentation with data combined with, we hope, an equal dose of deductive reasoning, it seems appropriate to discuss the data and methodology of the study in this section. For a study of corporate cash and security holdings, quarterly or monthly time series observations for a number of different firms would be desirable. Unfortunately, annual observations were all that could be obtained. Nevertheless, the particular sample of firms used has a number of advantages including eleven observations for each firm. The sample was collected and the accounts standardized by the Board of Governors of the Federal Reserve System.[1] The sample consists of balance sheet and income statement information for 11 years, 1946 to 1956, for each of 209 firms.[2]

1. A discussion of the data is given in the Federal Reserve Bulletin (June, 1956), p. 580.

2. The industries represented by the firms are food, tobacco, rubber,

For the purpose of this study the sample was split into a group of 44 firms chosen randomly, which will be called the subsample, and the remaining group of 165 firms, which will be referred to as the main sample. Summary data for the subsample and main sample are presented in Table 1, which shows the mean dollar value for 1947–1955 of cash, securities, sales and total assets for the firms. The firms in the main sample are split into four size groups by total assets. Ratios of the variables are also presented by size class and for all firms. The 209 firms studied accounted for about 1/6 of the cash, 1/5 of the marketable securities, and 1/6 of the sales of all United States non-financial corporations.

The purpose of using a subsample was to choose from several alternative relationships a form that could in turn be tested on the main sample.[1] Questions arise in empirical work as to the functional form in which to use variables and often neither theory nor conclusions from previous investigations are a helpful guide. Variables may be "deflated" by some size measure such as total assets or used in dollar terms. Theory does not tell which procedure is appropriate. Nor does it generally specify the way in which variables should be dated in relation to one another. A balance sheet item as of December 31, 1960, may in principle be related to a flow variable of 1960 or 1961 or some combination of these flows. Another question is the type of regression analysis to be used when a rectangular array of data such as the present sample is available. Is it necessary to run a separate regression for each firm or can firms be expected to share common values of some or all parameters?

Another advantage of working with a pilot group of observations is that, by reducing the number of relations examined for the main sample, it gives more meaning to relations tested on the main sample. Because a large number of regression equations were estimated using the data of the subsample, some of the relations could appear good by chance rather than because a good formulation of the behavior of firms had been found. The test of whether a good fit on the pilot sample is due to a good description of behavior or to chance must employ an independent group of observations such as those provided by the main sample. The process of hypothesis testing does not end here. As will be indicated, some of the formulations

---

petroleum, chemicals, iron and steel, non-ferrous metals, automobiles, other transportation equipment, machinery including electrical, and retail trade. Also available but not used in this study are data for some 50 firms in electric power and railroads.

1. The use of a subsample when the original sample is large may also economize on computing facilities. In the present case, where the subsample is 1/5 of the total, one could, with equal computing time, examine at least 10 relations on the subsample and test the seemingly best relation on the remainder of the sample or estimate three different relationships on all the observations.

TABLE 1

1947–1955 Averages of Cash, Securities, Sales, and Total Assets and Ratios for the Sample of 209 Firms (Dollar Figures in Millions)

	Main Sample						
	Asset Size Groups						
	10–70	70–180	180–500	500 and up	Total	Subsample	Total
Number of Firms	41	42	39	43	165	44	209
(1) Average Cash Holdings	4.19	9.66	23.40	57.30	23.96	26.68	24.54
(2) Average Securities Holdings	2.19	6.44	17.60	69.10	24.36	23.44	24.15
(3) Average Sales	74.7	134.2	372.2	862.2	365.3	530.5	400.1
(4) Average Total Assets	33.8	92.2	219.0	746.5	275.3	296.4	282.0
(5) Cash to Sales Ratio (1)/(3)	.056	.072	.063	.066	.066	.050	.061
(6) Cash to Assets Ratio (1)/(4)	.124	.105	.107	.077	.087	.090	.087
(7) Securities to Sales Ratio (2)/(3)	.029	.048	.047	.080	.067	.044	.060
(8) Securities to Assets Ratio (2)/(4)	.065	.070	.081	.093	.089	.079	.081
(9) Sales to Assets Ratio (3)/(4)	2.21	1.46	1.70	1.15	1.33	1.79	1.42

suggested by the subsample did not prove significant on the main sample and in turn the main sample also suggested further hypotheses. The principal conclusions from the subsample are incorporated in the theoretical discussion of Section 2 and the models of Sections 3 and 4.

2. Some Variables Influencing Corporate Cash and Security Holdings

The variables discussed in this section are by no means all those that will affect the cash and security holdings of firms. The discussion is mainly concerned with short-run influences on cash and securities of an individual firm. Variables that may cause cash and security holdings to differ among firms are largely neglected in the treatment. Nor will the treatment of short-run influences on a firm's cash and security holdings attempt to be exhaustive. The first part of this section discusses the level of transactions and the bill rate, while the second part treats the interrelationships among cash and securities and the other current account items.

A. The Level of Transactions and the Interest Rate

Interest rates and the volume of transactions have frequently been used to explain cash holdings and, to some extent, security holdings, and there seems no reason to break with tradition. In particular there have been a number of analytical and empirical studies of the relation of cash holdings to the volume of transactions, though the exact form of the relation has not generally been established. [1]

---

1. An extensive bibliography of empirical studies on the velocity of money is contained in a recent study by Richard T. Seldon, "Monetary Velocity in the United States," in Studies in the Quantity Theory of Money, Milton Friedman, ed., (Chicago: University of Chicago Press, 1955), pp. 179-195. A discussion of the relation of velocity to interest rates is contained on pp. 195-199.

William Baumol and James Tobin have argued analytically that cash held for transactions purposes will rise less than proportionately to the volume of transactions. Tobin emphasizes that other short-term assets are suitable for transactions balances, and both Baumol and Tobin emphasize the inverse relation that may be expected between the proportion of cash in transactions balances and the interest rate. See Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," Quarterly Journal of Economics, November, 1952, and Tobin, "The Interest Elasticity of the Transactions Demand for Cash," Review of Economics and Statistics, August, 1956.

If the relation between cash holdings and transactions is non-proportional, then estimation of a relation that is additive in the logarithms of the variables will often provide a better fit than estimation of a relation that is additive in the variables. Avram Kisselgoff, for aggregated data on a sample of corporations from 1921 to 1939, has used a relation between corporate cash and transactions that was additive in the variables. See "Liquidity Preference of Large Manufacturing Corporation (1921-1939)," Econometrica, October, 1945.

The relation between money and interest rates has also received extensive theoretical investigation; however, this literature will not be discussed here. The general view taken in this study is that corporations hold cash because the transactions costs into and out of short-term assets exceed the return on these assets for the period they will be held.[1] The cost of purchasing government securities is quite small, varying from about 5¢ per $1,000 to 30¢ per $1,000 depending on the thickness of the security market and the size of the purchase.[2] Suppose a firm receives $1 million today and must make a $1 million payment in a week. Even if the costs of buying and selling a government security earning 4% were as high as 60¢ (30¢ in and 30¢ out) per $1,000, there would be an excess of return over costs for holding $1 million of that security for a week.

The most important short-term asset that firms buy is Treasury bills. The tax anticipation certificate is a security especially suited to absorbing one type of excess funds of firms. In addition, repurchase agreements with government bond dealers and commercial paper of sales finance companies are especially tailored to the maturity needs of corporate lenders. Large

---

More recently, Friedman has estimated an aggregate relation between cash and income using an equation that was additive in the logarithms of the variables. See "The Demand for Money: Some Theoretical and Empirical Results," Journal of Political Economy, August, 1959. For the economy as a whole cash held for all purposes appears to have risen more than proportionally to income.

1. It is recognized that firms may hold demand deposits as compensating balances with banks from which they customarily borrow. Although no attempt is made to incorporate compensating balances into the analysis, the possible effect of these balances on our results is discussed in Section 5.

2. Broker's charges begin at $2.50 per $1,000 for government securities and decrease fairly quickly with the size of the purchase. When the size of purchase is $1 million of bills (or somewhat less for bonds) it is possible to deal directly with government bond dealers. The implicit transactions costs are the difference between bid and ask yield (for bills) or price (for bonds) of the government security dealers. This may vary according to the state of the money market and the length to maturity of the security. In terms of the purchase price of a bill the difference will frequently run 5¢ to 10¢ per $1,000. For other securities the difference may be as high as 1/32 of one percent of the price, or roughly 30¢ per $1,000. Contrary to the example above, the implicit transactions cost includes both purchase and sale of securities. For a more detailed discussion of the buying and selling of securities by dealers see Carl H. Madden, The Money Side of "the Street," (New York: Federal Reserve Bank of New York, 1959), pp. 47–72.

firms are able to earn interest on temporary excess funds for periods as short as a few days by use of either of these arrangements.[1]

Empirical studies employing aggregative time series data have found an inverse relation between cash holdings and the interest rate.[2] Critics of

---

1. Repurchase agreements are described by Hyman Minsky, "The Central Bank and Money Market Changes," Quarterly Journal of Economics, (May, 1957), pp.176-178, and by Madden, op. cit., pp. 61-62. Madden also describes the selling of their own commercial paper by sales finance companies. He says, "While most sales finance companies still market their notes through dealers, nine of the largest now sell them directly to investors, and in doing so have fashioned a new market for themselves. The dealer function—locating buyers and arranging purchase terms—is performed within the finance company organization. In effect the companies make a general offer to sell their IOU's at the various rates of discount, in any desired denomination, for any number of days from 5 to 270, and sometimes more. The notes are thus completely tailor-made for the customer." (p.69).

2. Kisselgoff, op. cit., related the interest rate to aggregate cash holdings of a sample of corporations for the period 1921 to 1939. He also found a strong inverse relation between "free" cash and the interest rate. Free cash is defined as current cash holdings minus the product of current cash holdings and the ratio of cash to transactions in 1929, a year when the cash-transactions ratio was at a minimum for the period. This procedure in effect assumes that free cash in 1929 was zero. Kisselgoff used some of the data of a study of businesses by the National Bureau of Economic Research. One of the monographs from this study, Corporate Cash Balances, 1914-1941, (New York: National Bureau of Economic Research, 1945), Chapters 4 and 5, by Friedrich Lutz, also employs the concept of free cash and his results support an inverse relation between free cash and interest rates. James Tobin examined the relation between the prime rate and free cash (calculated in a similar manner) for the economy as a whole during the pre-World War II period and found an inverse relationship. See his "Liquidity Preference and Monetary Policy," Review of Economics and Statistics, (May, 1947), pp. 130-131. Andrew Stedry, "A Note on Interest Rates and the Demand for Money," Review of Economics and Statistics, (August, 1959), pp. 303-307 has extended Tobin's computations through 1958 and has demonstrated the need for considering wealth in any long-run analysis of the demand for money. L. S. Ritter, "Income Velocity and Monetary Policy," American Economic Review, (March, 1959), pp. 120-129, has examined quarterly cash and free cash holdings for the economy as a whole (third quarter of 1957 represents the low point of the ratio of money to real income for the period) for the period 1948-1957, finding an inverse relation of cash holdings to the bill rate. Similar results were found by Henry A. Latané, "Cash Balances and the Interest Rate—A Pragmatic Approach," Review of Economics and Statistics, (November, 1954), pp. 456-460. Phillip Cagan

these studies have underlined some of the statistical problems of isolating the effects of interest rates on cash holdings but for aggregative data an inverse relationship appears well established. A statistical difficulty with time series analysis of the effects of interest rates is the typical positive correlation between income and interest rates or the existence of similar trends in income, cash and interest rates.[1] While it is likely that there will be some positive relationship between interest rates and the level of transactions of individual firms, it is doubtful that the relations will be as strong as for aggregative data. In addition, interest rates and the level of transactions are probably interdependent in the aggregate, and this need not be the case for the individual firm.[2] While there are likely to be estimating difficulties for micro-data not present in macro-data, an objective of this study is to estimate the effect of interest rates on money holdings in a way that avoids many aggregate time series problems.

The models of cash and security holdings developed below are theoretically concerned with the interest rate on Treasury bills, tax certificates, commercial paper, repurchase agreements, other short-term securities, and short-term bank loans. Unfortunately, the short-term rates earned and paid are not available for the individual firms of the sample. For any particular asset or liability, it is therefore necessary to use the same rate for all firms in each year, and because the rates on short-term assets and liabilities will be highly correlated, a choice must be made as to which

---

presents related results, "The Demand for Currency Relative to the Total Money Supply," Journal of Political Economy, (August, 1958), pp. 303-329, finding that the interest rate on demand and time deposits has been an important determinant of the ratio of currency to total demand and time deposits that the public wishes to hold.

1. The study of Kisselgoff, op. cit., was criticized by Acheson J. Duncan, "'Free Money' of Large Manufacturing Corporations and the Rate of Interest," Econometrica, (July, 1946), p. 251, because of a similar trend of both free cash and interest rates in the data that Kisselgoff used. Clark Warburton, "Monetary Velocity and the Rate of Interest," Review of Economics and Statistics, (August, 1950), pp. 256-257, has similarly argued that the results of Tobin's study (see previous footnote) could be interpreted in terms of the correlation of unrelated trends rather than a functional relationship between money and interest rates.

2. Latané, op. cit., explicitly treats this problem in his study. He regresses the reciprocal of the rate of interest (1/r) on the ratio of money to income (M/Y), and also regresses (M/Y) on (1/r). He uses an average of these regression lines as an approximation of the true relationship among these variables.

rate to use. The bill rate has been chosen because a major concern of this study is the substitution of cash for short-term securities, and a major portion of these securities are Treasury bills. As the bill rate is correlated with other short-term rates, there will not be much harm in using the bill rate as "the interest rate" in applications where it may be that other rates would be more relevant.

B. The Relationship of Cash and Securities to Other Current Accounts

The current accounts of firms consist of three principal liabilities, accounts payable, taxes due, and short-term bank loans, and four principal current assets, cash, securities, accounts receivable, and inventories. There are other current assets and liabilities that for any particular firm or group of firms may be more important than some of the above, but we will limit the present discussion to the above accounts.

Three of the accounts mentioned are unlikely to be influenced by the levels of cash and security holdings. These are taxes due, accounts receivable, and accounts payable. Taxes due are in effect a short-term interest free loan of the government to firms making profits. Although firms may always pay their tax liability before it is due, there is no reason for them to do so. The firms should always let the taxes due liability be as large as legally possible, since these funds can always earn the interest rate on tax certificates, and at worst earn zero as cash.[1] Accounts receivable should also be unaffected by the levels of cash and securities. The terms on accounts receivable are set by the firm and are stable in the short-run. If, for example, the interest rate changes, it is not business custom to alter the terms on their accounts receivable.[2] This means that in the short-run a firm's receivables are largely determined by the customers of the firm. A firm's customers may be expected to increase their use of receivables as they increase their purchases from firms and as the rate of interest rises.

The relation between receivables and the interest rate occurs because of the stability of the terms on receivables over the cycle.[3] Firms

---

1. Firms may hold these funds as cash in the form of compensating balances.

2. Part of the firms in the present sample are engaged in retail trade and it will be true that some retailers alter the terms on their accounts receivable in relation to business conditions. In particular, installment credit terms are subject to cyclical change, though firms frequently do not finance this credit themselves. While the statement in the text needs to be qualified to take account of installment credit, the fact that a few firms in our sample do alter their terms on receivables over the cycle will not be very important to the subsequent analysis.

3. Though terms vary widely among firms, probably the most frequent terms on accounts receivable are a 2% discount for cash payment by the tenth day of the month after purchase. Other terms on accounts receivable

in effect offer a given number of days credit at a zero interest rate no matter what the cost of borrowing funds. When the costs of borrowing rise, this should induce customers who have not charged their purchases to do so. Those customers who have for convenience paid invoices early upon their arrival are led, by the higher costs of cash, to consider delaying payment until the tenth of the month. Some customers might not take the cash discount if paying for purchases by the tenth of the month requires securing a new bank loan.[1] These actions by customers in response to interest rate changes may cause quite a large increase in accounts receivable.[2]

It is possible to test whether the customers of large firms, and large firms as customers, increase their use of trade credit as the interest rate

---

are presented by Charles W. Gerstenberg, Financial Organization and Management of Business, revised edition, (New York: Prentice-Hall, 1934), p. 458. Industries where the delivery date is somewhat uncertain often give longer terms on receivables. The striking example is capital goods producers whose turnover of receivables before World War II was almost twice as long as most other industries. The rate of turnover of receivables is given for various industries by Walter Chudson, Pattern of Corporate Financial Structure, (New York: National Bureau of Economic Research, 1945), pp. 11-43, and Albert Koch, Financing of Large Corporations, 1920-1939, (New York: National Bureau of Economic Research, 1943), pp. 53-56.

1. Charles Silberman, in "The Strange Money Shortage," Fortune, (March, 1957), p. 123, says, "Big companies have had to let their receivables rise for several reasons. Many small firms, for example, have stopped discounting their bills. Why pay in ten days to claim a flat 2 per cent discount, many treasurers reason, if they have to borrow from the banks in order to do so? Similarly, larger companies that still discount their bills now delay payments until the tenth day (or whenever the discount runs out), rather than pay as soon as the invoice is received."

2. If the terms of receivables are such that a purchase on the first of a month can be delayed to the tenth of the next month, then the maximum length of turnover of receivables is roughly 40 days, assuming the cash discount is taken. Thus the turnover of receivables can range from zero days (no one charges) to 40 days, or from zero to roughly 1/9 of annual sales. Denoting receivables by R, annual sales by S, and the average turnover of receivables by T, then $R/S = T/365$. If T moves from 28 to 30 days for a firm with $500 million in sales, then receivables will increase by $3.7 million. Silberman reports that, in 1956 and 1957, firms found their receivables lengthening considerably. Continental Can, for example, found its T moving from 26 to 28 days in a year, requiring nearly a $10 million increase in receivables, while other firms reported 10% or larger increases in T in a year. See Silberman, "The Strange Money Shortage," Fortune, (March, 1957), p. 258.

rises. Given the level of sales it is expected that the large firms of this sample will find their receivables rising with the interest rate. Also, given the level of purchases of the firms of this sample, it is expected that their accounts payable, i.e., the credit they take as customers, will rise with the interest rate. The results of these tests are reported in Section 3.

Why should firms leave the terms on accounts receivable stable during a cycle? The reason is that the return on increased sales is so high relative to costs of borrowing for large firms that they are quite willing to allow their customers to take whatever credit they wish. If the return from increased sales is high, then the return on inventory investment, if the alternative is lost sales, will also be very high when sales are rising. Since non-financial corporations are selling goods, and the return from increased sales is likely to be high, relative to borrowing costs and also to the bill rate, for large firms, they are likely to give priority to short-run adjustments in inventories. That is, when there is a sales increase, firms will first attempt to adjust their inventories to a higher level and only then worry about adjusting cash and securities to the new level of sales. If this is a reasonable description of firm behavior, then inventories may be assumed to be unaffected in the short-run by the levels of cash and security holdings and the effect of inventories on cash and security holdings may be examined without simultaneously considering the influence of cash and securities on inventories. [1]

The assumption that net receivables and inventories are not influenced in the short-run by the levels of cash and security holdings allows us to use net receivables and inventories as independent variables in regression equations explaining cash and security holdings. An indirect test of this assumption is incorporated in the models used to explain cash and security holdings. These models test whether firms adjust their cash and security holdings to their desired levels immediately or with a lag. If cash and securities are adjusted to desired levels only with a lag, it is evidence consistent with the assumption that short-run inventory adjustments of the firm are given priority in time to adjustments of cash and securities.

Having put the cart before the horse, we can turn now to the more interesting question of why inventories and net receivables might enter into an explanation of cash and security holdings. When a firm's sales increase,

---

1. This statement may be amended slightly. Once firms have at least partially adjusted their inventories to a higher level of sales, and firms are beginning to adjust their cash and securities, there may well be some feedback from the cash and security position to the inventory position. Such feedback means that inventories are affected by the levels of cash and securities. However, so long as the influence on inventories of cash and securities comes at a different time from the influence of inventories on cash and securities, it is not necessary to consider these accounts simultaneously.

receivables rise, and the return on inventories, inventory investment, and accounts payable also rise. Since we are here considering large firms, the impact of a sales increase on net receivables is decidedly positive.[1] Thus during a business expansion a source of finance is needed for the increase in net receivables and inventories. One obvious and traditional source of such finance is short-term bank loans. As an alternative to short-term bank loans, firms could borrow long-term and hold funds as securities when sales were low and as inventories and net receivables when sales were high.

It is a hypothesis of this study that firms will find it economical to substitute between the use of securities and short-term bank loans in the finance of these periodic increases in net receivables and inventories. The reason it may be economical for firms to use securities is that the rates on short-term bank loans during the peak of a cycle may be high relative to the long-term rate at the time when firms borrow long-term. The use of securities will also depend on the character of fluctuations in inventories and net receivables, as well as the costs of long-term and short-term borrowing and the bill rate. This hypothesis is consistent with the empirical observation that firms predominantly use long-term sources of finance for their stock of net receivables and inventories.[2]

It is expected, then, that inventories and net receivables will be inversely related to security holdings, and that the causation is one way running from inventories and net receivables to securities. It is also expected that cash holdings will bear the same relation to these variables and that both cash and security holdings will be substitutes for short-term bank loans. Since security holdings of firms do earn a return and can be liquidated without capital loss at fairly short notice, it is expected that the above relations will be much stronger for securities than cash.

These are the relationships between cash and securities and the other current accounts that are expected over a firm's inventory cycle. As the receipts of firms are not particularly stable, it is important that they be held constant if the above relationships are to be observed. Thus if a firm has an unexpectedly large inflow of receipts during an inventory expansion it may be in a position to expand inventories and, in addition, augment its cash and security holdings. If a measure of unexpected receipts is not included as a variable in the model, we would observe in the above example a positive relation between inventories and cash and security holdings rather

---

1. In the present sample, on the average, receivables were 13.9% of total assets, payables 6.7%, and hence net receivables 7.2% of total assets. The average annual amount of net receivables for all 209 firms is about $4.3 billion.

2. The average of the ratio of receivables plus inventories to total assets of the 165 firms is .3886, while the average of short-term bank loans to total assets is .0237. As short-term bank loans are only 6.1% of the stock of net receivables and inventories, it is clear that large firms prefer or find more readily available other sources of finance than short-term loans.

than the negative relation that is expected. The variable measuring unanticipated receipts should be positively related to cash and security holdings and negatively related to short-term bank loans.

The reasons that firms may temporarily put receipts into cash and securities are: a) there are no liabilities that may be immediately retired whose borrowing costs are as great as the return on bills (e.g., there is no point in using unexpected receipts to reduce accounts payable since, assuming the cash discount is taken, this liability costs less than the return on bills), and b) the firm has not had time to put money into other more lucrative assets. In short, the firm may get "caught" with an unexpected influx of funds which it must hold in a form yielding a relatively low return. As usual it is argued that the rational firm would hold as much as possible of such receipts in the form of securities rather than cash.

The discussion of this part may be summarized as four propositions about the current accounts of firms. Propositions (1) and (3) are not directly tested and may be considered hypotheses of the study.

1) Short-run fluctuations in accounts payable, accounts receivable, and taxes of firms will not be influenced by the level of cash and security holdings.

2) When interest rates rise, accounts receivable may be expected to rise with the level of interest rates for the same reasons that firms will increase their use of accounts payable, namely, the terms on accounts receivable are stable in the short-run.

3) The inventory position gets first priority in short-run adjustments of firms and any influence of securities and cash on the desired inventory holdings will come at a different point in time from the influence of inventories on cash and securities. Thus, these accounts do not have to be considered as simultaneously determined in the short-run.

4) Securities and, to a lesser extent, cash will be used as a substitute for short-term bank loans in the financing of short-period increases in inventories and net receivables.

This leaves us with an explanation of cash and security holdings involving some of the other current assets and liabilities, the level of transactions, and the interest rate. There are other variables that might be considered, but we have introduced only one, a measure of unexpected receipts. It is felt that such a variable is important because in the short-run unexpected receipts are apt to appear as cash or securities whether or not the firm wishes additional cash and securities. In Section 3 tests are made of propositions (2) and (4). In Section 4 lagged adjustment models of cash and security holdings are estimated. These estimates should provide some evidence about the inventory assumption, and test the relation of cash and securities and the explanatory variables.

3. Estimates of Equations for Bank Loans and Trade Credit

A. Notation and General Description of Tests

This section is divided into three parts. The notation and definitions of variables that are used in all of the empirical work are presented

in this part. In the second part we report the results of estimating a canonical correlation, which attempts to test in a very general way the validity of the short-run adjustment model presented in the previous section. The third part reports the results of testing the hypothesis that customers of firms increase their use of accounts receivable as the interest rate rises.

Definitions of Variables: All Variables Are a Ratio to Average Total Assets of the Firm, except the Bill Rate.

Notation	Definition of Variable
T	Expected level of transactions of a firm in year t, measured by current sales.
B	The rate of interest. It is the 90 day treasury bill rate, average of October, November, and December rates in year t.
C	Cash holdings of firms, including demand deposits, time deposits (negligible) and currency.
S	All fixed claim marketable securities. Equities are not included.
D	Accounts payable of firm.
L	Short-term bank loans of maturity under a year.
R	Accounts receivable.
I	Inventories.
P	Retained earnings, i.e., profits after taxes minus cash dividends.
N	Taxes due, a balance sheet item.
"i"	The subscript "i" on a variable is an index running over the number of firms, and for this sample i = 1, 2, ..., 165.
"t"	The subscript "t" on a variable is an index of time. t = 48, 49, ...56, i.e., there are 9 observations on each variable for each firm. A variable subscripted "t-1" will be observations on the variable from 1947 through 1955.
"	The symbol " indicates an estimate.
Δ	Denotes a one year change in a variable, e.g., for a stock, $\Delta C_{it} = (C_{it} - C_{it-1})$, so ΔC_{455} is the change in the stock of cash between Dec. 31, 1954, and Dec. 31, 1955, for the 4th firm in the sample. Similarly for flow variables, ΔP_{13249} is retained earnings for the year ended Dec. 31, 1949, minus retained earnings for the year ended Dec. 31, 1948, for the 132nd firm in the sample.

A few comments should be made about the measures of variables used.

Sales is used as a measure of transactions because there appears little difference between the two variables in annual data. It is possible to derive a measure of transactions from balance sheet and income data that includes all but interannual purchases and sales of assets. When such a measure of transactions is computed from annual data it is very nearly the same as sales.[1] In the remainder of this study the terms "transactions" and "sales" will be used interchangeably.

The balance sheet information on securities does not provide a breakdown by maturity or type of security. Included in the measure of securities are corporate and government bonds, tax certificates, bills, notes, and commercial paper. A survey by Fortune magazine of 276 large corporations indicates that typically the securities held were short-term Treasury obligations.[2]

Finally, all variables for each firm have been divided by a size measure that is constant for each firm over all years but differs among firms. The size measure used was average total assets for each firm for the period 1947–1955. This procedure seems advisable because, in cross section data for firms, the variance of the dependent variable among firms is frequently related to the size of the independent variable. The residual term in a linear relationship among variables such as we will be using is likely to be larger for larger values of the independent variable; this violates the assumption of homoskedasticity underlying estimates of the standard errors of regression coefficients.[3] For firm data stated in dollar magnitudes, the size of most independent variables is related to the size

---

1. For the subsample of firms a concept of transactions was used that measured outpayments associated with current production. The correlation (R^2) between the level of sales and the level of transactions was .982 and for changes in sales and changes in transactions R^2 was .905. When sales and the measure of transactions were used in regressions the coefficients were virtually the same. A more inclusive concept of transactions was used by Lutz and Kisselgoff and is described in Lutz, op. cit., Appendix B. Kisselgoff, op. cit., p. 335, who uses a part of the data used by Lutz, comments that there was little difference between this concept and sales.

2. Charles E. Silberman, "The Big Corporate Lenders," Fortune (August, 1956), pp. 111–114.

3. The observations on an individual firm over a short period of years may be considered as a sample of observations from the population of all observations that could have been observed for the firm during the period. The true error term in a regression equation estimated for a firm from this population is assumed to be unrelated to the independent variable. If the assumption of homoskedasticity was not met for each individual firm it would be appropriate to divide all variables by a size variable that differed between observations for each firm as well as differing among firms.

of the firm. Dividing all variables by a size measure is intended to reduce the size of the true error component of the estimated residuals.

In the following parts and Section 4 the least squares estimates of several regression equations will be presented. For each regression equation the estimates of the coefficients and their standard errors, R^2, and the F ratio for the equation will be presented. The standard errors of the regression coefficients allow us to test whether a coefficient is significantly different from any particular value. For all coefficients it is reported whether the estimated value is different from zero at the 5 and 1% levels of significance.[1]

R^2 is the fraction of variance of the dependent variable of a regression equation that is explained by the independent variables. Contrary to usual usage we will always use correlation and correlation coefficient to refer to the value of R^2. This should not lead to any ambiguity since R, which in conventional usage is the correlation coefficient, is not used in any of the discussion. To test whether R^2 is significantly different from zero we compare the "F" value of a regression equation with the critical value of F at the 5 and 1% levels.[2]

In the use of canonical correlation it is not possible to make the usual tests of significance, and our hypothesis has to be accepted or rejected on the basis of a sign test and to some extent intuitive judgment. Where tests of significance are used, we reject the null hypothesis that a coefficient is equal to zero only if the critical value of F or t we have obtained is greater than the critical value of F or t at the 1% level. Although we report results at the 5% level it is felt the 1% level is appropriate in this study because of the large number of observations.[3]

---

1. Because N is large for the present tests, the critical value of "t" (the ratio of a coefficient to its standard error) is virtually the same for all coefficients in all regression equations, namely 1.96 and 2.58 at the 5% and 1% levels respectively.

2. The "F" value of a regression equation is the ratio of the explained variance of the dependent variable divided by the degrees of freedom (minus 1) used in estimating the regression equation to the unexplained variance divided by the remaining degrees of freedom.

3. The computations were carried out at the Yale University Computing Center, and the availability of a computer greatly eased the task. The calculations were further facilitated by the availability of computer programs. I am particularly grateful to Harold W. Watts and Donald D. Hester of the Cowles Foundation for Research in Economics at Yale University. Watts is to be thanked for his extremely useful set of programs for estimating regression equations, and Hester for his program for canonical correlation.

B. Canonical Correlation Analysis.[2]

Proposition 4 of Section 2 considered the interrelationships of cash, securities, and short-term bank loans, and net receivables, inventories, and unanticipated receipts. One set of equations that could be used to analyse these relationships is

$$(3\text{-}1)\quad \Delta C_t = \alpha_{11} + \alpha_{12}\,\Delta I_t + \alpha_{13}\,\Delta(R\text{-}D)_t + \alpha_{14}\Delta(P\text{-}N)_t + u_{1t}$$

$$(3\text{-}2)\quad \Delta S_t = \alpha_{21} + \alpha_{22}\,\Delta I_t + \alpha_{23}\,\Delta(R\text{-}D)_t + \alpha_{24}\Delta(P\text{-}N)_t + u_{2t}$$

$$(3\text{-}3)\quad \Delta L_t = \alpha_{31} + \alpha_{32}\,\Delta I_t + \alpha_{33}\,\Delta(R\text{-}D)_t + \alpha_{34}\Delta(P\text{-}N)_t + u_{3t}$$

where u_1, u_2, and u_3 are residual or error terms, and the α's are parameters. The change in retained earnings plus the change in taxes owed is used in this set of equations as a measure of unanticipated receipts in any year. As we will discuss in Section 4 there is no particular reason to suppose that $\Delta(P+N)$ is either a good or bad approximation of unanticipated receipts.

These three equations do not allow us to explore the relationships among securities, cash, and bank loans. We could, it is true, specify a system of equations that would let us take account of the interrelationships among them. A simpler approach which answers our questions about C, S, and L is to use the method of canonical correlation.

We define the variable "q," where $q_t = \beta_1\Delta C_t + \beta_2\Delta S_t + \beta_3\Delta L_t$, and where the β's are weights, and the variable "r," where $r_t = b_1\Delta I_t + b_2\Delta(R\text{-}D)_t + b3\Delta(P\text{-}N)_t$, and the b's are weights. The canonical weights are those values of the β's and b's which maximize (or give the canonical) correlation between q and r.[2] We expect that the canonical weights of inventories, net receivables, and short-term bank loans will have the same sign, and those of cash, securities, and unexpected receipts will all have the opposite sign. This pattern of signs is consistent with the substitution of short-term bank loans for cash and securities in the financing of inventories and net receivables, and one test with the canonical correlation is a sign test.

Also some subjective judgment about the hypothesis can be made from examining the correlation coefficients. Let us denote the squared canonical correlation coefficient as $\underline{R}^2$ and denote the largest R^2 occurring in the three regression equations, (3-1), (3-2), and (3-3), as R^2_m. $\underline{R}^2$ must always be greater than or equal to R^2_m. If $\underline{R}^2$ is not much larger

---

1. Canonical correlation theory has been described by Harold Hotelling, "Relations Between Two Sets of Variates," Biometrika, vol. 28 (1936), pp. 321-377.

2. The canonical weights are invariant up to a linear transformation and the computing method used makes use of this property.

than R_m^2 then the hypothesis may be rejected. If the sign test is passed, and $\underline{R}^2$ is considerably larger than R_m^2, then the hypothesis is accepted. Because the distribution of the differences between $\underline{R}^2$ and R_m^2 is not known, it is not possible to test for a significant difference in these coefficients, and only rough judgments can be formed.

The results of estimating the regression equations and the canonical correlation are presented in Table 2. The "F" values are all significant at the 1% level which means that the three regression equations each explain a significant amount of the variance of the dependent variables. The regression coefficients are all significantly different from zero at the 1% level except net receivables for cash and all are of the expected sign.

Examination of the signs of the b's and β's from the canonical correlation show that net receivables, inventories, and short-term bank loans have negative signs, and cash, securities, and unexpected receipts positive signs. Thus when inventories and net receivables rise, short-term bank loans go up, and cash and securities go down. As this is the expected pattern of signs the sign test is consistent with the conclusion that cash and securities are substitutes for short-term bank loans in the financing of inventories and net receivables. The squared canonical correlation coefficient of .440 exceeds by 15% the value of R_m^2, which was .290 in equation (3-3). The difference between these coefficients appears substantial and is interpreted as evidence that firms prefer to use combinations of securities, cash, and short-term bank loans for short-term finance, rather than any one of these sources alone.

What of the importance of cash, securities, and bank loans relative to each other in the financing of inventories and net receivables? The canonical weights, the β's, tell us that short-term bank loans are still used more relative to securities and cash, as short-term bank loans have the largest weight. This conclusion is important because it means that large firms do have short-term credit requirements for which they are apparently dependent on bank loans.

C. Relation of Accounts Payable and Net Receivables to the Bill Rate

In the course of the discussion of the behavior of accounts receivable it was suggested that large firms as customers will tend to increase their use of trade credit as the interest rate rises. The test of this proposition performed here is fairly simple, and should not be considered an attempt to specify a complete explanation of accounts payable. It is expected that, given the level of purchases from other firms, a firm would increase its accounts payable as the interest rate rises. This hypothesis may be expressed as

$$(3\text{-}4)\quad D_t^* = \alpha_{41} + \alpha_{42}(T + \Delta I)_t + \alpha_{43}B_t + u_{4t}$$

where D_t^* is desired accounts payable, and u_{4t} is the error or residual term. We do not have data on purchases by the firm, but the level of sales plus the change in inventories should be a good approximation. The assumption is made that $D_t = D_t^*$, since we see no reason why, given the terms on which firms can obtain accounts payable, the observed volume of accounts payable of a firm

TABLE 2

Estimated Regression Equations and Canonical Correlation

Variables	Regression Equation (3-1)	Regression Equation (3-2)	Regression Equation (3-3)	Canonical Correlation and Canonical Weights***
Intercept	.0025	.0060	-.0050	
	(.0009)*	(.0012)*	(.0011)*	
Change in Cash (ΔC)	1.0			1.000
Change in Securities (ΔS)		1.0		1.132
Change in Short-term Bank Loans (ΔL)			1.0	-1.460
Change in Inventories	-.0518	-.1783	.3084	-.705
	(.0134)*	(.0182)*	(.0159)*	
Change in Net Receivables [Δ(R-D)]	-.0511	-.1728	.3159	-.709
	(.0208)**	(.0282)*	(.0247)*	
Change in Retained Earnings and Taxes [Δ(P+N)]	.1978	.3830	-.1303	.824
	(.0162)*	(.0221)*	(.0193)*	
R^2	.093*	.198*	.290*	
F	50.64	121.67	201.19	
Degrees of Freedom	1481	1481	1481	
Squared Canonical Correlation Coefficient				.440

* Significantly different from zero at 1% level.

** Significantly different from zero at 5% level.

*** These weights are based on a weight of 1.000 for cash. Because of the computational method used, these weights will differ by an uninteresting factor of proportionality from the canonical weights.

should be different than the amount it desires.

With the substitution of D_t for D_t^* into (3-4) we have the equation we wish to estimate except for one further modification. The equation to be estimated is

(3-5) $D_{it} = \alpha_{41i} + \alpha_{42}(T + \Delta I)_{it} + \alpha_{43} B_t + u_{4it}$,

where u_{4it} is the error or residual term. The subscript "i" on α_{41} indicates that separate intercepts are specified for each firm. The lack of subscript "i" on the interest rate indicates it is a constant for all firms in a particular year. The theoretical argument for specifying separate intercepts for each firm is the same as will be made in the next section with respect to equations for cash and securities. Discussion of this point will be deferred until then.

The estimates of (3-5) are

$$(3\text{-}5)'' \quad D_{it} = \alpha''_{41i} + \underset{(.0016)}{.0450}\,(T + \Delta I)_{it} + \underset{(.0980)}{.4961}B_t + u''_{4it},$$

where " indicates an estimate. R^2 for the regression equation is .832. The critical value of F at the 1% level is 1.33, while the F value for the equation is 39.31.[1] Both of the regression coefficients are significantly different from zero at the 1% level.[2] Our hypothesis that, given the level of purchases, firms increase their use of accounts payable when the interest rate rises is consistent with these results. We turn now to the behavior of the customers of large firms.

Rather than work directly with accounts receivable, net receivables will be examined as a dependent variable in order to find out if large firms on net supply credit to small firms when the interest rate rises. Specifying for the individual firm

$$(3\text{-}6) \quad (R\text{-}D)^*_t = \alpha_{51} + \alpha_{52}T_t + \alpha_{53}B_t + u_{5t},$$

and assuming $(R\text{-}D)_t = (R\text{-}D)^*_t$, and specifying separate intercepts we have

$$(3\text{-}7) \quad (R\text{-}D)_{it} = \alpha_{51i} + \alpha_{52}T_{it} + \alpha_{53}B_t + u_{5it}.$$

While we have specified accounts payable to be related to purchases, there seems no reason to expect net accounts receivable to be related to purchases. Rather, since accounts receivable will be related very closely to sales, it seemed appropriate to use sales in (3-6). The estimate of (3-7) is

$$(3\text{-}7)'' \quad (R\text{-}D)_{it} = \alpha''_{51i} + \underset{(.0040)}{.0289}T_{it} + \underset{(.2415)}{2.1124}B_t + u''_{5it}.$$

R^2 is .771 and F is 26.72, the equations and the two slope coefficients being significant at the 1% level.

The estimated coefficients for net receivables are quite large, and there is some reason to expect this may be due to a significant trend inn net

1. The number of observations is 1485, the number of parameters estimated is 167 (165 intercepts plus 2 slope coefficients), so the number of degrees of freedom in the numerator of the F ratio is 166, and in the denominator 1318.

2. It is not necessary to estimate the α_{41i}'s in order to estimate the other coefficients and since there was no particular interest in the α_{41i}'s we did not estimate them. In computing the other coefficients we make use of the fact that the least squares estimate of $\alpha_{41i} = \overline{D}_i = \alpha_{42}(\overline{S + \Delta I})_i - \alpha_{43}\overline{B}$. Substituting this expression for α_{41i} in (3-4) gives us an alternative form of our original regression equation. This alternative form is much simpler to compute when the number of separate intercepts is quite large.

receivables. While accounts payable and accounts receivable may individually have some trend, cyclical movements, especially in a nine year time series, are apt to be equally important. However, even if receivables and payables display strong cyclical movements, their difference may still have a large trend over a nine year period.

To test this conjecture the sum of the estimated residuals ($\Sigma u''_{5it}$) was calculated for each year for all firms. Examination of the series of these sums of residuals suggested that time might be included in our regression equation for net receivables. The resulting estimates are

$$(3\text{-}8)'' \quad (R\text{-}D)_{it} = \alpha''_{61i} + \underset{(.0043)}{.0182T_{it}} + \underset{(.3040)}{.9795B_t} + \underset{(.0008)}{.0049_t} + u''_{6it}$$

where "t" is now also a variable, u''_{6it} is our estimated residual term in this equation, and α''_{61i} the separate intercepts. The F ratio is 27.48, and R^2 is .777 and significant. Although the sales and interest coefficients are still significant at the 1% level, they are considerably reduced by the inclusion of time. The introduction of time adds virtually nothing to the explained variance of the regression equations, and the significance of the time coefficient results directly from the reduced magnitude of the sales and interest coefficients.

In Table 3 the sums of residuals in each year for the accounts payable regression equation and for the two net receivables equations are presented. Also in Table 3 are the sums of the original variables by year. A casual glance at column (2) suggests that the residuals for the accounts payable equation display a fairly random pattern. However, time was added to the accounts payable equation, anyway, and the results were as expected. The coefficient of time was not significantly different from zero and the coefficients of sales and the bill rate were identical in the 4th decimal place in the two equations.

The sums of residuals ($\Sigma u''_{5it}$) for each year for equation $(3\text{-}7)''$ for receivables are presented in column (4), while the sums of residuals ($\Sigma u''_{6it}$) after time has been introduced are shown in column (5). It appears that the pattern of the residuals in column (5) is more random than those in column (4) and it is concluded that the introduction of time has improved the explanation of net receivables.

It should be pointed out that a good part of the explained variance of the regression equations for accounts payable and net receivables is due to the estimation of separate intercepts. If the variance of D_t and (R-D) is measured around the means of these variables for each firm, the values of R^2 (partial coefficients of determination) that measure only the explained variance due to the independent variables may be calculated. This measure of R^2 for accounts payable, when B_t and $(S + \Delta I)_t$ are the independent variables, is .509. For net receivables, when B_t, S_t, and time are the independent variables, R^2 due to these variables is .184.[1] For the sake of

1. It may be useful to define these measures of correlation more precisely. For the regression equation for accounts payable,

TABLE 3

Analysis of Residuals for Accounts Payable and Net Receivables by Year

Year	Sum of Accounts Payable* (1)	Sum of Residuals for Accounts Payable (2)	Sum of Net Receivables* (3)	Sum of Net Receivable Residuals** (4)	Sum of Net Receivable Residuals (5)
1948	-2.9858	-.3918	-4.9840	-1.7713	-.0250
1949	-3.8164	-.2676	-5.4166	-1.5483	-.9039
1950	-1.2662	.8206	-2.2610	.0592	.6316
1951	.0376	-.0979	-.4234	-.2719	.4744
1952	1.0250	.2330	.7893	-.3647	.1902
1953	.4974	-.5081	.9161	.7470	-.1033
1954	.0906	.2870	.9236	2.7704	.1043
1955	2.2445	-.0233	3.9461	.4257	-.2694
1956	4.1733	-.0579	6.5100	-.0961	-.0989

* These figures are the sums of the variable as deviations from the mean of the variable for all the observations.

** The residuals in column (4) are estimated without the inclusion of time as a variable, while those in column (5) are estimated when time is included in the regression equation.

$$R^2 = [1 - \frac{\Sigma\Sigma\, u''^2_{4it}}{\Sigma\Sigma (D_{it} - \bar{D})^2}],$$

when the variance of D is calculated about the mean of all observations. The second measure of correlation discussed above is

$$R^2 = [1 - \frac{\Sigma\Sigma\, u''^2_{4it}}{\Sigma\Sigma (D_{it} - \bar{D}_i)^2}],$$

where the variance of D is calculated about the mean for each firm. In Appendix 2 [p.161 and equation (A-3)], it is shown why the second R^2 might be considered appropriate for measuring the explained variance due to the estimated slope coefficients on the independent variables. The choice of the term "partial coefficient of determination" may not have been felicitous.

completeness, two other regression equations for accounts payable and net receivables have been estimated.

These two equations do not assume that $D_t = D_t^*$, and $(R-D)_t = (R-D)_t^*$, but rather the following relationships are specified for each individual firm,

$$(3-9)\quad \Delta D_t = B_1 (D_t^* - D_{t-1}) + u_{7t}$$

$$(3-10)\quad \Delta(R-D)_t = B_2 [(R-D)_t^* - (R-D)_{t-1}] + u_{8t}$$

Substituting the expressions for D_t^* from (3-4) into (3-9) and (3-6) for $(R-D)_t^*$ in (3-10), and adding separate intercepts we obtain

$$(3-11)\quad \Delta D_{it} = B_1\underline{\alpha}_{41i} + B_1\underline{\alpha}_{42}(T + \Delta I) + B_2\underline{\alpha}_{43}B_t + B_2 [-D_{it-1}) + v_{1it}$$

$$(3-12)\quad \Delta(R-D)_{it} = B_2\underline{\alpha}_{51i} + B_2\underline{\alpha}_{52}T_{it} + B_2\underline{\alpha}_{53}B_t + B_2 [-(R-D)_{it-1}] + v_{2it}$$

where $v_{1it} = (B_1 u_{4it} + u_{7it})$ and $v_{2it} = (B_2 u_{5it} + u_{8it})$, and where again common slopes are assumed for all firms, but separate intercepts are specified. As the α's from estimating (3-11) and (3-12) may differ from (3-5) and (3-7), they are underscored in (3-11) and (3-12).

The estimated regression equation for accounts payable is

$$(3-11)''\quad \Delta D_{it} = \underline{\alpha}''_{41i} + \underset{(.0016)}{.0324}\ (T + \Delta I)_{it} + \underset{(.0943)}{.7991}B_t + \underset{(.0247)}{.7003}\ (-D_{it-1}) + v''_{1it}$$

and for net receivables is

$$(3-12)''\quad \Delta(R-D)_{it} = \underline{\alpha}''_{51i} + \underset{(.0031)}{.0106}T_{it} + \underset{(.1871)}{1.1011}B_t + \underset{(.0227)}{.3078}\ [-(R-D)_{it-1}] + v''_{2it}$$

The regression coefficients in both equations are significant at the 1% level and the F ratios indicate that a significant amount of the variance of the dependent variable is explained in both equations.[1] R^2 for accounts payable is .442 and R^2 not including the explained variance due to the separate intercepts is .411. For net receivables the corresponding correlations are .266 and .134.

The coefficients in (3-11)" and (3-12)" are fairly comparable in magnitude with (3-5)" and (3-8)", and the conclusions we wish drawn from these estimates will not depend on which equations are considered. Since it is felt that observed and planned, or expected, accounts payable and net receivables should be equal, equations (3-5)" and (3-8)" will be used as the basis for our conclusions.

The results suggest that even though large firms are induced to increase the credit they take from other firms in response to changes in the interest rate, their customers are even more responsive. We would expect

---

1. When time was introduced into the equation for net receivables the "t" value of the time coefficient was 1.02, so time has not been included above.

the interest coefficient for net receivables to be zero if the customers of large firms showed the same response as do large firms as customers. Quite evidently this is not the case.

The magnitude of the interest coefficient on net receivables indicates that large firms supply a sizeable credit stream to other firms, governments, and consumers when interest rates rise. The coefficient of .9795 tells us that a rise in the interest rate of 1% (say from 2% to 3%) increases net receivables as a ratio to total assets by nearly 1%. For the sample, this means roughly a $450 million increase in net receivables for a rise of 1% in the bill rate.[1] These results are discussed further in Section 5.

4. Estimates of Regression Equations for Cash and Securities.

In Part A of this section the models of cash and security holdings developed in Section 2 are expressed as regression equations. The estimates of these regression equations are given in Part B. One set of regression equations developed in Part A specifies separate intercepts for each firm, and these intercepts have been analysed by industry and asset size of the firm in Appendix 1. Part C of this section reports the results of analysing the error terms from certain equations for cash and securities. A complete description of the tests used on the residuals is given in Appendix 2.

A. The Regression Models of Cash and Securities

The discussion of Sections 1 and 2 suggest the following relations:

$$(4\text{-}1)\quad C_t^* = \alpha_{60} + \alpha_{61}T_t + \alpha_{62}B_t + \alpha_{63}I_t + \alpha_{64}(R - D)_t + v_{3t},$$

$$(4\text{-}2)\quad S_t^* = \alpha_{70} + \alpha_{71}T_t + \alpha_{72}B_t + \alpha_{73}I_t + \alpha_{74}(R - D)_t + v_{4t},$$

where C_t^* and S_t^* are respectively the desired levels of cash and securities of an individual firm, the α's are parameters and the v's are residual terms. In addition to the traditional variables, the level of transactions and the bill rate, net receivables, and inventories are included in accord with the discussion of Section 2.

Will $C_t = C_t^*$ and $S_t = S_t^*$? In the discussion of Section 2 it was suggested that short-run adjustments in inventories and net receivables will take precedence over adjustments of cash and securities. Another proposition of Section 2 was that firms may have to hold unexpected receipts in the form of cash and securities whether or not they wish to increase these holdings. If this argument is correct then the observed levels of C_t and S_t are unlikely to be their desired levels, and another approach is necessary.

The following identities are used:

$$(4\text{-}3)\quad \Delta C_t = \Delta C_t^p + \Delta C_t^u, \text{ and}$$

1. Average total assets of the firms of this sample are $275.3 million, so [165 x 275.3 x .009795] = $445 million.

$$(4\text{-}4)\quad \Delta S_t = \Delta S_t^p + \Delta S_t^u,$$

where the superscript "p" denotes a planned change, and the superscript "u" an unplanned change. Planned changes in cash and securities are written as some proportion of desired changes:

$$(4\text{-}5)\quad \Delta C_t^p = \beta_3 (C_t^* - C_{t-1}),$$

$$(4\text{-}6)\quad \Delta S_t^p = \beta_4 (S_t^* - S_{t-1}).$$

The coefficient β_3 tells us what portion of a desired change in cash is made in a year, and similarly for β_4 for securities. It is expected that the β's lie between 0 and 1.0. If $\beta_3 = .5$, it is interpreted to mean that because of the many balance sheet adjustments that must be made, a firm can only adjust its cash holdings one-half of the way to their desired level in any year.

Finally unexpected receipts are introduced into the model:

$$(4\text{-}7)\quad \Delta C_t^u = \alpha_{65} \Delta N_t + v_{5t},$$

$$(4\text{-}8)\quad \Delta S_t^u = \alpha_{75} \Delta N_t + v_{6t},$$

where the v's are residual terms. The change in taxes owed is used as the measure of unexpected receipts in the above formulation. It may be recalled that in Section 3 $\Delta(P + N)$ was used. Since the decision as to which of these variables to use as a measure of unanticipated receipts is non-theoretical, discussion of this point is postponed until the next part of this section.

Substituting equation (4-1) into (4-5), and equations (4-5) and (4-7) into (4-3), equation (4-9) is obtained.

$$(4\text{-}9)\quad \Delta C_t = (\beta_3\alpha_{60}) + (\beta_3\alpha_{61})T_t + (\beta_3\alpha_{62})B_t + (\beta_3\alpha_{63})I_t + (\beta_3\alpha_{64})(R\text{-}D)_t + \beta_3(-C_{t-1}) + \alpha_{65}\Delta N_t + w_{1t},$$

where $w_{1t} = (\beta_3 v_{3t} + v_{5t})$. Substituting equation (4-2) into (4-6), and equations (4-6) and (4-8) into (4-4), equation (4-10) is obtained.

$$(4\text{-}10)\quad \Delta S_t = (\beta_4\alpha_{70}) + (\beta_4\alpha_{71})T_t + (\beta_4\alpha_{72})B_t + (\beta_4\alpha_{73})I_t + (\beta_4\alpha_{74})(R\text{-}D)_t + \beta_4(-S_{t-1}) + \alpha_{75}\Delta N_t + w_{2t},$$

where $w_{2t} = (\beta_4 v_{4t} + v_{6t})$. Parentheses are put around pairs of coefficients, as $(\beta_4\alpha_{71})$, in these equations to indicate that only the product of the two coefficients can be estimated directly.

Can these two relations be estimated pooling the data for all firms? It is assumed here that the coefficients on the independent variables will tend to share common values. However, work with the subsample of firms suggested that firms frequently may not have common

intercepts. That is, $(\beta_3\alpha_{60})$ and $(\beta_4\alpha_{70})$ are likely to differ among firms. Examination of the cash equation may illustrate the problem. Consider the coefficient β_3 on lagged cash $(-C_{t-1})$. If C_{t-1} is more cash than the firm desires this year it is expected that the change in cash will be negative. Suppose for a particular firm $C_{t-1} = .10$ this year which was more cash than the firm desired so it decreased its cash holdings. Suppose for some other firm with the same values of T, I, (R-D), B, and ΔN, it is observed that $C_{t-1} = .10$. Would it be expected that the second firm would also decrease its cash holdings by the same amount this year, assuming the value of β_3 was the same for both firms? This would be expected to occur only if factors affecting differences in average cash holdings between the two firms were effectively explained by T, I, (R-D), B, and ΔN. It is not believed that these other variables do adequately explain differences in average cash holdings among firms. In addition, work with the subsample of firms was not very suggestive as to other variables that might be related to inter-firm differences in mean cash holdings. While it would be desirable to specify a model that explained differences in the average values of variables among firms, if this is not easily done, why not abstract from these differences by the use of separate intercepts for each firm?

This discussion may also be illustrated graphically. In Figure 1 the marginal relation, given the effects of the specified independent variables, between ΔC_t and C_{t-1} is plotted for two firms A and B. The slope, β_3, is assumed the same for both firms, and the mean values of changes in cash may or may not be equal, though they have been assumed to be zero for both firms in Figure 1. The only difference between the two firms is the mean of $(-C_{t-1})$. If common slopes and intercepts were estimated pooling the data for the two firms, the slope would be some relation like the dotted line in Figure 1, which could differ markedly from the "true" value of β_3. If separate intercepts of OA for firm A and OB for firm B were estimated, the common slope estimate would be β_3, which illustrates why it is desirable to estimate separate intercepts if the mean values of the variables differ in a manner not easily explained.

If separate intercepts are estimated by the method of least squares the slope β_3 is estimated from deviations of ΔC and $(-C_{t-1})$ from the respective means of these variables for each firm. This is illustrated in Figure 2 where the mean of ΔC and $(-C_{t-1})$ for firm A have been removed from A's observations, and similarly for firm B. The same type of reasoning used here is the basis for the estimation of separate intercepts for accounts payable and net receivables in Section 3. Also it may be recalled that in Section 3, common intercepts were employed in the regression equations estimated in Part B of that section. As all the variables entering these regression equations were changes in stocks and flows, rather than stocks or flows, there is no reason to expect that differences in the means of the variables are not explained by the other variables in the equations.

The addition of separate intercepts to (4-9) and (4-10) gives

Figure 1

Illustration of the Use of Separate Intercepts

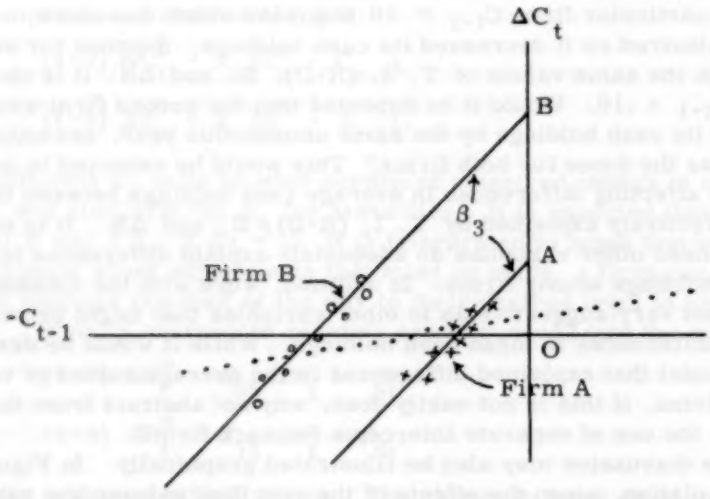
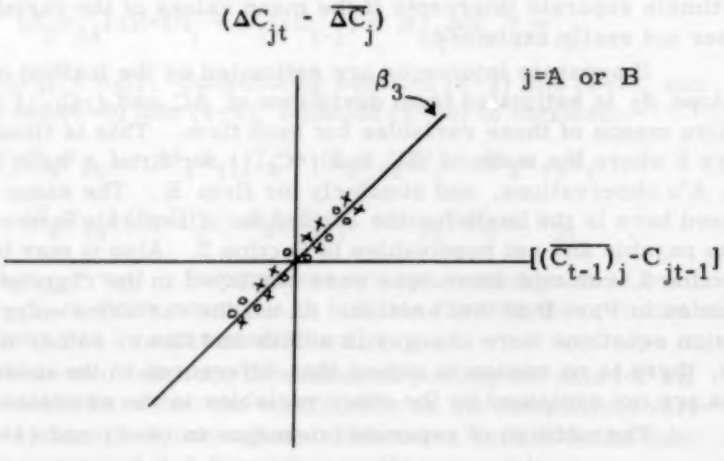


Figure 2

Transformation of Variables When Separate Intercepts Are Used



$$(4\text{-}11)\quad \Delta C_{it} = (\beta_3\alpha_{60})_i + (\beta_3\alpha_{61})T_{it} + (\beta_3\alpha_{62})B_t + (\beta_3\alpha_{63})I_{it} + (\beta_3\alpha_{64})(R\text{-}D)_{it} + \beta_3(-C_{it-1}) + \alpha_{65}\Delta N_{it} + w_{1it},$$ and

$$(4\text{-}12)\quad \Delta S_{it} = (\beta_4\alpha_{70})_i + (\beta_4\alpha_{71})T_{it} + (\beta_4\alpha_{72})B_t + (\beta_4\alpha_{73})I_{it} + (\beta_4\alpha_{74})(R\text{-}D)_{it} + \beta_4(-S_{it-1}) + \alpha_{75}\Delta N_{it} + w_{2it}.$$

Both (4-9) and (4-11) for cash and (4-10) and (4-12) for securities will be estimated to see whether the specification of separate intercepts does help in the explanation of changes in cash and securities.

B. The Estimates of Regression Equations for Cash and Securities.

The estimates of the two cash equations and the two securities equations are presented in Table 4. In Table 4 equation A for cash is (4-9) of the previous part, and equation B is (4-11). Similarly for securities, equation A in Table 4 is (4-10) of the previous part, equation B is (4-12). An F test is used to determine whether an equation specifying separate intercepts provides a better explanation of changes in cash than does an equation with a common intercept. The addition to explained variance due to the separate intercepts divided by the number of additional intercepts is the numerator. The denominator is the residual variance of the equation with the additional intercepts divided by the remaining degrees of freedom. The F ratio for cash is thus

$$F = \frac{\frac{1.555416 - 1.043537}{164}}{\frac{1.043537}{1314}} = 3.94.$$

With 164 and 1314 degrees of freedom the critical value of F at the 1% level is 1.33, so the hypothesis that separate intercepts add significantly to the explanation of changes in cash is accepted. The test is identical for securities where $F = \frac{\frac{3.066702 - 2.311055}{164}}{\frac{2.311055}{1314}} = 2.62$. For securities, the hypothesis that specification of separate intercepts adds significantly to the explanation of changes in securities is also accepted at the 1% level.[1]

1. An assumption of the above test is that the slope coefficients on the independent variables are the same for each firm. The assumption of common slopes has not been tested here. To test for common slopes equations (4-9) for cash and (4-10) for securities would be estimated for each firm separately and the residual variance for the individual regression equations summed over all firms would be compared with the residual variance from a pooled regression. The reason for not making the test of common slopes is that it is a substantial investment in computation that would yield regression equations for individual firms that one would probably not otherwise wish to estimate. That is, the regressions for each firm would only have two degrees of freedom each which is apt to produce from sampling error

TABLE 4

Estimates of Regression Equations for Cash and Securities***
(Standard Errors of Regression Coefficients in Parentheses)

Independent Variables	Cash Equation A. (4-9)	Cash Equation B. (4-11)	Securities Equation A. (4-10)	Securities Equation B. (4-12)
Lagged Cash ($-C_{it-1}$)	.1672 (.0150)*	.6312 (.0239)*		
Lagged Securities ($-S_{t-1}$)			.1235 (.0140)*	.4782 (.0231)*
Sales (T)	.0043 (.0008)*	.0333 (.0029)*	.0017 (.0011)	.0263 (.0042)*
Bill Rate (B)	-.0235 (.1405)	-.3848 (.1384)*	-.2834 (.1996)	.0008 (.2000)
Inventories (I)	-.0228 (.0079)*	-.0391 (.0120)*	-.0367 (.0076)*	-.1321 (.0179)*
Net Receivables (R-D)	-.0029 (.0053)	-.0098 (.0153)	-.0307 (.0101)*	-.0260 (.0228)
Change in Taxes (ΔN)	.2199 (.0229)*	.1207 (.0206)*	.4721 (.0322)*	.3237 (.0310)*
R^{2**}	.140	.423	.191	.390
F	40.15	5.67	58.19	4.95
Degrees of Freedom	1478	1314	1478	1314
Sum of Squared Residuals	1.555416	1.043537	3.066702	2.311055

* Different from zero at the 1% level of significance.

** The variance of ΔC is computed about $\Delta\overline{C}_i$, the mean of ΔC for each firm. R^2 for regression equation B above is .392 and significant. For securities the corresponding R^2 for equation B is .366.

*** The intercept term for equation A for cash is .0144, and for equation A for securities is .0318, both coefficients being significant at the 1% level.

This is not quite the final word on the subject of separate intercepts. It is shown in Appendix 2 that the same slope coefficients of the independent variables in the cash and securities equations would result if the independent variables were expressed as deviations from their mean for each firm, and the change in cash and the change in securities were expressed as deviations from the mean of all observations.[1] This formulation uses only one intercept and yields values of R^2 of .371 for cash and .352 for securities, and patterns of significance identical with equations B of Table 4 for cash and securities.[2] It does not really make any difference to the conclusions of the text whether the set of equations just described, or equations B of Table 4 are used, and so we have chosen to work with the latter set because it is slightly less cumbersome to describe. However, in the analysis of residual variance it is not convenient to use the equations with separate intercepts. A complete description of the above set of equations and the analysis of residual variance is contained in Appendix 2.

For cash, all of the coefficients in equation B in Table 4 are of the expected sign, and all but the coefficient on net receivables are significant at the 1% level. For securities, the coefficients in equation B of Table 4 are of the expected sign, and all but the coefficients on net receivables and the bill rate are significant at the 1% level. One problem in estimating the effect of the bill rate on cash and securities is that we only have the quantities of net receivables and inventories, and not the returns on these assets. Since substitution of inventories for cash and securities is due to the increased return on inventories, the estimates of the effects of the bill rate on cash and securities will be affected by the use of the quantity of inventories. The problem is further complicated by the fact that the bill rate, the return on inventories, and the quantity of inventories are correlated. We cannot really say how the estimates of the bill rate coefficients are

---

some wide but uninteresting variations in coefficients. Since the common slope assumption has not been tested, the above F tests could have the interpretation that has been placed on them or they could have the interpretation that enough of the differences in slope coefficients among firms has shown up as differences in intercepts to allow separate intercepts to add significantly to the explained variance of cash and securities.

1. The reasons for specifying such a set of equations is implicit in the discussion of separate intercepts in Part A of this section. The reason it is not necessary to actually use separate intercepts for the cash and security equations is that the dependent variables are changes in stocks, and the variations in the means of changes in stocks among firms over a short period is not expected to be, or is not in fact, large.

2. This statement means that the same variables have significant "t" values in the two sets of equations. The standard errors of the coefficients differ in the two sets of equations, and in particular they are larger for equations B of Table 4.

affected because we do not know how accurately the quantity of inventories mirrors the return on inventories. In any event it must be concluded that the present estimates yield no significant relation between the bill rate and securities.

The fact that net receivables is not a significant variable in equations B for cash and securities is not too surprising. It is expected that inventories and net receivables would have the same general relationship with respect to all variables but the bill rate. While it was expected that net receivables would be positively related to the bill rate, it turns out that inventories in the sample are also related positively to the bill rate evidently because it has not been possible to hold the rate of return on inventories constant. If there is not much difference in the behavior of two independent variables entering into a regression equation, it may be difficult to distinguish their effects on the dependent variable. The reason for treating the variables separately was to see if cash and securities are used more to finance one or the other of the assets. The answer is either that cash and securities are used predominantly for inventories, or that the model is not sensitive enough to the differences between receivables and inventories to allow an answer to the question. Equations B for cash and securities are not much changed if net receivables is dropped as a variable, or if net receivables is added to inventories as a composite variable, so in analysing the results in Section 5, the original equations (4-11) and (4-12) will be used.

In Sections 2 and 3, three different regression equations involving the relation of cash and securities to net receivables have been reported. In none of these regression equations for cash was net receivables a significant variable at the 1% level. This suggests that, though a negative relation between these variables is found persistently, the relation is not strong enough to take into account. Again, this conclusion is subject to the reservation above, namely, that the models do not allow us to distinguish very sharply between the effects of inventories and net receivables. For securities, only regression equation B of Table 4, of the three equations examined, yields a relation between net receivables and securities that is not significant at the 1% level. It is concluded that the negative relation found between securities and net receivables is strong enough to take into account, but that equation (4-12) for securities does not allow a good estimate of this relation.

Changes in taxes due was used as the measure of unanticipated receipts in the regression equations in this section. Four variables had been considered. These were the change in sales, change in profits, change in taxes, and the change in taxes plus profits. In the work with the subsample it appeared that an average of lagged sales was a better measure of expected transactions, the conceptual variable to which firms would adjust their cash and security holdings, than current sales.[1] When this lagged measure of

1. Two other variables that appeared to have explanatory value for the subsample, but did not for the main sample, should be mentioned. The first was investment. It was reasoned that if firms were planning large capital

sales was used for transactions, the change of sales was used to measure unexpected receipts. For the main sample it turned out that current sales worked better than a lagged average of sales, and because the change of sales is highly correlated with current sales, the change of sales was no longer considered as a measure of unexpected receipts.[1]

The other three measures of unexpected receipts are not much different in the results they produce. In general it was better to use either the change in taxes or the change in profits plus taxes. Taxes are important because the change in the tax law in 1954 made it necessary for firms to pay one-half of their tax liabilities in the year they accrued and one-half in the following year, instead of all in the following year. This reduced the size of the permissible taxes owed account for any given level of profits, and Silberman reports that this had an appreciable effect on the working capital of firms in 1955 and 1956.[2] Since profits would not reflect this "unanticipated" drop in receipts for firms, it seemed desirable to use taxes and profits, or taxes alone. In general ($\Delta P + \Delta N$) did about as much better for cash as (ΔN) did better for securities, where the difference in R^2's are in the third decimal place. There did not seem any justification for taking the best of both

---

expenditures they might build-up funds as securities, and perhaps also as cash. A fairly strong negative relation between investment and changes in securities was found for the subsample. While the relation of securities and investment was consistently negative for the main sample, the relation was not generally significant at the 1% level. The second variable was a lagged measure of the liquidity position of the firm. The liquidity position was measured as total fixed claim liabilities minus current assets. It was expected that large values of this liquidity measure would be associated with a subsequent build-up of cash and securities. However, no persistent relationship was found either for cash or securities in the main sample. One reason why these variables might not be important in the model used is that the lagged measure of cash and securities may bury the effects of such variables. Thus a large value of lagged securities means securities decrease next year, whether or not the large value of lagged securities is due to a build-up for investment. Similarly a low value of lagged cash will result in a build-up in cash next year, whether or not the low value of lagged cash was a reflection of a low liquidity position of the firm.

1. By worked better it is meant that for identical regression equations for cash and securities, the coefficients and their "t" values for current sales were larger than for the lagged average of sales. Two measures of lagged sales were used. The first was $.5(T_{t-1}+T_{t-2})$, and the second was $.25(T_t+T_{t-2}+2T_{t-1})$.

2. Charles E. Silberman, "The Strange Money Shortage," Fortune (March, 1957), p. 123.

worlds, so the change in taxes was used in both regression equations (4-11) and (4-12), as indicated above.[1]

C. Discussion of the Residuals for Cash and Securities

The regression equations estimated in this study pool observations over industries, asset size groups, and years. The advantage of pooling the data is that it allows us to make more general statements. Since the models used here are of a very simple form and involve some misspecifications of relationships among the independent variables, the results will be of an inexact character no matter what level of aggregation is used. While the inexact nature of the results may justify our procedure, pooling of data is unlikely to be without costs. These costs may be assessed to some extent by examining the error terms from the regression equations.

Residuals have been estimated for cash and securities from equations with the same slope coefficients as those in equations B of Table 4. The residuals have been classified by year, industry, asset size, and by the size of certain independent variables. There was no significant difference in the means of residuals for cash by year using the homogeneity of means test at either the 5% or 1% level of significance. For securities there was a significant difference in the means by year at either the 5% or 1% level of significance. With only nine annual observations it is difficult to discern cyclical or other patterns in any time series including the present residuals. From examination of the means of the residuals and the independent variables we observe no obvious systematic factors that would explain the apparent differences in the means of residuals of securities by year, and believe the differences are due to circumstances in particular years. There is no trend in the residuals of cash and securities at the 1% level, though there is a trend in the residuals for securities at the 5% level. It is suggested that any importance of time in the explanation of securities is due to the observations for 1948, but other judgments are certainly possible, and the interested reader may wish to consult Appendix 2, and especially Table (A-3).

Classification of the residual variance by the size of sales, inventories, and the lagged values of cash and securities, revealed that there is heteroskedasticity for both cash and securities with respect to some of the independent variables. It appears that dividing variables by total assets may have reduced one cause of heteroskedasticity only to introduce another.

---

1. In the regression equations in Part B of Section 3, $(\Delta P + \Delta N)$ was used as the measure of unexpected receipts. This was considered consistent procedure in the sense that the measure chosen was better for securities which was the criterion above. (ΔN was better in these regressions for cash and short-term bank loans.) It would also have been consistent to use the change in taxes in all regressions. Differences in R^2's between ΔL, ΔS, and ΔC, and ΔN and $(\Delta P + \Delta N)$ are in the third decimal place so that our consistency, or lack of it, will have had little effect on the results.

This analysis also suggested that non-linear forms might have been more appropriate for describing certain relationships than the linear forms used here.

The "explained variance" for subgroups of observations, due to the regression equations estimated over all observations, has been examined. The subgroups consist of observations classed by year, industry, and asset size. For each of these subgroups except 1943 for securities and 1956 for cash the hypothesis that the "explained variance" is equal to zero is rejected at the 1% level. The quotation marks about "explained variance" indicates that the present usage differs from the usual usage. All of these terms, tests, and conclusions are discussed in more detail in Appendix 2.

5. The Results: Analysis, Summary, and Implications for Monetary Policy

A. Analysis of the Estimated Equations for Cash and Securities

In this part of the section, the coefficients for cash and securities estimated in Section 4 will be discussed in more detail. The estimated regression equation for cash is

$$(4-11)'' \quad \Delta C''_{it} = (\beta_3 \alpha_{60})''_i + .0330T_{it} - .3838B_t - .0391I_{it} - .0098(R-D)_{it} + .6312(-C_{it-1}) + .1207\Delta N_{it} + w''_{lit} \ldots$$

This is a short-run relationship. Setting ΔC_{it} and ΔN_{it} equal to zero and substituting C^*_{it} for C_{it-1}, an estimate of "equilibrium cash holdings" is obtained, which is

$$(4-1)'' \quad C^{*''}_{it} = \alpha''_{60i} + .0528T_{it} - .6096\,B_t - .0619\,I_{it} - .0155(R-D)_{it} + \frac{w_{1it}}{\beta''_3} .$$

By identical procedure an estimate of the equation (4-2) of equilibrium security holdings may be obtained.

The derivation of these equilibrium coefficients whether done as above, or directly from the equations of Section 4, is simple enough, and their interpretation for sales and the bill rate is straightforward. Because the firm is not able to adjust its cash balances to their desired levels immediately, the observed adjustment will be less than the long-run adjustment. To estimate the long-run adjustment we must divide the short-run coefficient by the speed of adjustment coefficient. However, because it is not possible to compartmentalize cash and security holdings, we must assume, for example, that the full adjustment of security holdings to sales takes the same time as the adjustment of security holdings to inventories. Actually the latter adjustments are apt to be very nearly immediate and the equilibrium coefficients for inventories and net receivables are for this reason probably too high, and no conclusions are drawn from the values of these coefficients.

The original coefficients will be referred to as short-run coefficients, and the derived coefficients as long-run or equilibrium coefficients. Both

sets of coefficients are presented with their standard errors in Table 5. The standard errors of the derived coefficients are only approximate.[1] There has been no change in the pattern of significance between the short-run and long-run coefficients. The coefficients in Table 1 show the change in cash and securities as a ratio to total assets for a unit change in the independent variable. The coefficients may be easier to interpret if they are expressed as elasticities as in columns (3) and (4) of Table 5. These elasticities are computed at the mean values of the variables, except that for the variables which are changes, as ΔC, the mean value of C is used, rather than the mean of the change in the variable.[2] The mean values of the variables are presented in column (5).

Elasticities of linear relations are frequently not useful because they depend on the initial position chosen. However, the initial values used here, the mean values of the variables for the sample, may provide a reasonable position from which to consider small departures. The estimated elasticities in Table 5 tell us the short-run and long-run percentage change in the dependent variable for a 1% change in the independent variable, given that the

---

1. Using the notation of equation (4-11) of the previous section, p. 143, the approximation used is

$$s^2_{\alpha_{6j}} = \frac{s^2_{(\beta_3\alpha_{6j})}}{\beta_3^2} + \frac{(\beta_3\alpha_{6j})^2}{\beta_3^4}\, s^2_{\beta_3} - \frac{2(\beta_3\alpha_{6j})}{\beta_3^3}\, \mathrm{cov}[(\beta_3\alpha_{6j}),\beta_3)],$$

where j = 1, 2, 3, and 4, for the four long-run slope coefficients, and where s^2 denotes sample variances, and where the variance and covariance terms on the right side are obtained from the covariance matrix of coefficients. This approximation of the variance of a ratio from the parameters of the components of the ratio is given by G. Udny Yule, and M. G. Kendall, An Introduction to the Theory of Statistics, 13th ed., rev., (London: Charles Griffin and Company, 1947), formulas 16.8 and 16.9, pp. 299-300. The approximation is suitable for the present case because the speed of adjustment coefficients are expected to be between zero and one, and, importantly, are not expected to take on values very close to zero.

2. The computation of the short-run elasticity of cash with respect to sales may be illustrated. The elasticity formula is $\Delta C/C \div \Delta T/T$, where in this instance, ΔC is the coefficient of sales, and C is the mean of cash, so the numerator is .0333/.1042 = .3196. The coefficient of T for cash is for a unit change in T, so $\Delta T = 1.0$, and T is the mean of T, so the denominator is 1/1.7465, and the resulting elasticity is .558. The reason for not using the mean of changes in cash is that for most purposes one is interested in the elasticity of cash with respect to sales, not the elasticity of changes in cash with respect to sales.

TABLE 5

Short and Long-Run Coefficients and Elasticities for Cash and Securities
(Standard Errors of Coefficients in Parentheses)

Variable	Short-Run Coefficient	Long-Run Coefficient	Short-Run Elasticity	Long-Run Elasticity	Means of Variables
		Part A. Cash			
Lagged Cash ($-C_{t-1}$)	.6312 (.0239)*	1.0000	.6312		.1042
Sales (T)	.0333 (.0029)*	.0528 (.0059)*	.558	.885	1.7465
Bill Rate (B)	-.3848 (.1384)*	-.6096 (.2191)*	-.062	-.098	.0167
Inventories (I)	-.0391 (.0120)*	-.0619 (.0190)*	-.115	-.183	.3075
Net Receivables (R-D)	-.0098 (.0153)	-.0155 (.0242)	-.008	-.012	.0811
Change in Taxes (ΔN)	.1207 (.0206)*		.093		.0800**
		Part B. Securities			
Lagged Securities ($-S_{t-1}$)	.4782 (.0231)*	1.0000	.4782		.0734
Sales (T)	.0263 (.0042)*	.0550 (.0097)*	.626	1.309	1.7465
Bill Rate (B)	.0008 (.2000)	.0017 (.4338)			.0167
Inventories (I)	-.1321 (.0179)*	-.2762 (.0389)*	-.553	-1.156	.3075
Net Receivables (R-D)	-.0260 (.0228)	-.0544 (.0585)	-.029	-.061	.0811
Change in Taxes (ΔN)	.3237 (.0310)*		.353		.0800**

* Different from zero at the 1% level of significance.

** This is the mean of N as a ratio to total assets, not the mean of ΔN. In computing elasticities the means of C, S, and N are used rather than the means of changes in these variables.

other independent variables are at their mean values.[1]

Kisselgoff's findings, based on aggregated data for a sample of firms during the period 1921–1939 imply an elasticity of cash with respect to transactions of 1.14.[2] Since Kisselgoff's estimates assume that actual cash balances are desired balances it is probably appropriate to use the long-run elasticity of .885 of this study for comparison. If it is assumed the variance of the sample means of cash and sales for each of the two studies is zero, an estimate of the standard errors of these elasticities may be calculated.[3] They are .485 for Kisselgoff's elasticity and .051 for the equilibrium elasticity of this study. Since differences between the elasticities could be the result of sampling error, there does not seem much point in pursuing the comparison.[4]

It would appear that the estimated long-run sales elasticity here is less than one. This suggests that cash holdings may grow less than proportionately to sales. However, the cross-section statistics in Table 1 indicate that as the sales of firms get larger there is no clear variation in the ratio of cash to sales, which may mean that the sales coefficient for

---

1. The effects on the elasticities of other variables entering into the regression equations may be illustrated. The interest elasticity of cash will be less if sales are greater than the mean value. If the value of sales was greater than its mean value then, because of the positive coefficient on sales, the ratio of cash to total assets would be greater than its mean value. In this case any change in cash for a given change in the bill rate would be a smaller percentage change in cash than if sales was equal to its mean value. Thus the interest elasticity of cash would be inversely related to the value assumed for sales, and the elasticity of cash to sales will be positively related to the value assumed for the bill rate.

2. Kisselgoff, op. cit., pp. 336–338, presents the data underlying his estimates. The above elasticity was computed using his data and the estimated coefficient on transactions. This elasticity assumes that the interest rate is at its mean value for the period.

3. The elasticity of cash to sales (long-run) in these studies is $\overline{T}\alpha_{61} \div \overline{C}$. If $\overline{C}$ and $\overline{T}$ are assumed to have zero variance then the variance of the elasticity is $[(\overline{T}/\overline{C})^2 \quad s^2_{\alpha_{61}}]$ and the standard error of the elasticity estimates may be easily calculated.

4. If the elasticity estimates are normally distributed then with 95% probability Kisselgoff's elasticity of cash to sales would lie between .07 and 2.11. The 95% confidence limits for the elasticity estimate of this study are .78 to .99. As the limits of the present estimate are contained within the limits of Kisselgoff's elasticity it is concluded that there is no significant difference in the estimates. The usual test of equality of means is not appropriate here because the variances differ markedly.

cash is a little low.[1]

The elasticity of securities with respect to sales is greater than 1.0, suggesting the ratio of securities to sales rises with sales. This is in agreement with the pattern found in Table 1 where the ratio of securities to sales in the smallest asset size groups is .029 while for the largest it is .080.

Kisselgoff has also estimated the elasticity of cash with respect to the interest rate, but it is not very meaningful to compare his estimate with that of the present study.[2] A comparison of the coefficients shows that a 1% increase in the interest rate, say from 2% to 3%, results in an 11.3% decrease in cash ($\Delta C/C$), using Kisselgoff's estimates. For this study the short-run percentage decrease in cash is 3.7% and the long-run decrease is 5.9%. Assuming the mean of cash has a variance of zero the standard deviation of Kisselgoff's estimate is .030, and for the long-run coefficient of this study the standard error is .021.[3] As the present equilibrium estimate is over

1. As the interest coefficient on cash shows, corporations during the period of this study have been willing to hold less cash as interest rates rose. However, it is unlikely that the interest coefficient has picked up all of the long-run effects of generally rising interest rates during the period, e.g., the development of repurchase agreements between corporations and government bond dealers, and methods of speeding up payments and receipts. If the interest coefficient does not pick up these effects, then the coefficient on sales may be unduly low. This could be called a bias in the present study if the methods of economizing cash developed because of higher interest rates are not retained when, and if, interest rates decline to their early postwar levels. A description of some methods corporations have developed for economizing cash is given by Charles Silberman, "The Big Corporate Lenders," Fortune (August, 1956).

2. The reason is that Kisselgoff used the interest rate on two to five year government bonds while the bill rate is used here. Longer term interest rates have a higher mean (Kisselgoff's mean rate was 2.95% as compared with 1.67% for this study) and exhibit smaller fluctuations, in general, than short-term rates. This means that long-term interest rates, everything else the same, will produce higher interest elasticities. Kisselgoff's elasticity of cash with respect to the interest rate on two to five year bonds was -.332. Kisselgoff also estimated the elasticity of free cash with respect to the interest rate. Free cash was defined as observed cash in any year minus current sales times the ratio of cash to sales in 1929, a year when the cash sales ratio was at a minimum. The estimated elasticity was -1.26 (op. cit., p. 254). Acheson J. Duncan, op. cit., p. 251, has pointed out that there is a high probability that the close relation between free cash and the interest rate is spurious because the correlation of deviations from the trend of free cash and the interest rate was only -.03.

3. The long-run percentage decrease in cash due to a 1% rise in the bill rate is .01 $\alpha_{62} \div \overline{C}$, and the variance of this estimate if the variance of $\overline{C}$ is zero, is .0001 $s^2_{\alpha_{62}} \div \overline{C}^2$, from which the standard error is computed.

two standard deviations from Kisselgoff's estimate, it is likely that there is a significant difference in the coefficients.[1] While it has been suggested that Kisselgoff's estimate may have an upward bias, there is also some reason to suppose the present estimate may also have a bias, though not necessarily downward.[2]

The interest coefficients of Table 5 may be used to examine the proposition of Baumol and Tobin that the proportion of cash in transactions balances is inversely related to the interest rate.[3] The relation between transactions balances, cash, and the bill rate is illustrated in Figure 3. It is assumed that the sum of cash and securities are transactions balances, though as some amount of securities (and cash) is set aside as a buffer for changes in inventories and net receivables, this figure will be high. It is also assumed that the other variables in the model are held constant. The ratio of cash plus securities to total assets is plotted on the horizontal axis of Figure 3A, and the bill rate on the vertical axis.

As it is assumed in Figure 3 that all variables are initially at their mean values, one point on Figure 3A is the mean of cash plus securities to total assets and the mean of the bill rate. This point is .178 on the horizontal axis of Figure 3A and .0167 on the vertical axis. In Figure 3B the ratio of cash to transactions balances is on the horizontal axis, and the initial point is the mean of the bill rate and the mean of (C/C+S), which is .585. In both Figures 3A and 3B the short-run and long-run relations are plotted, the two relations intersecting at the initial points discussed above.

If the interest rate rises from its mean value of 1.67% to say 3%, then in Figure 3A the firm moves up the short-run relation curve to point a. If the interest rate stays at 3% for some time then the firm does not stay at point a, but rather moves over to point b on the long-run curve, since point b represents the full adjustment of the firm to a 3% interest rate. If after the bill rate had remained at 3% for some time, there was another

---

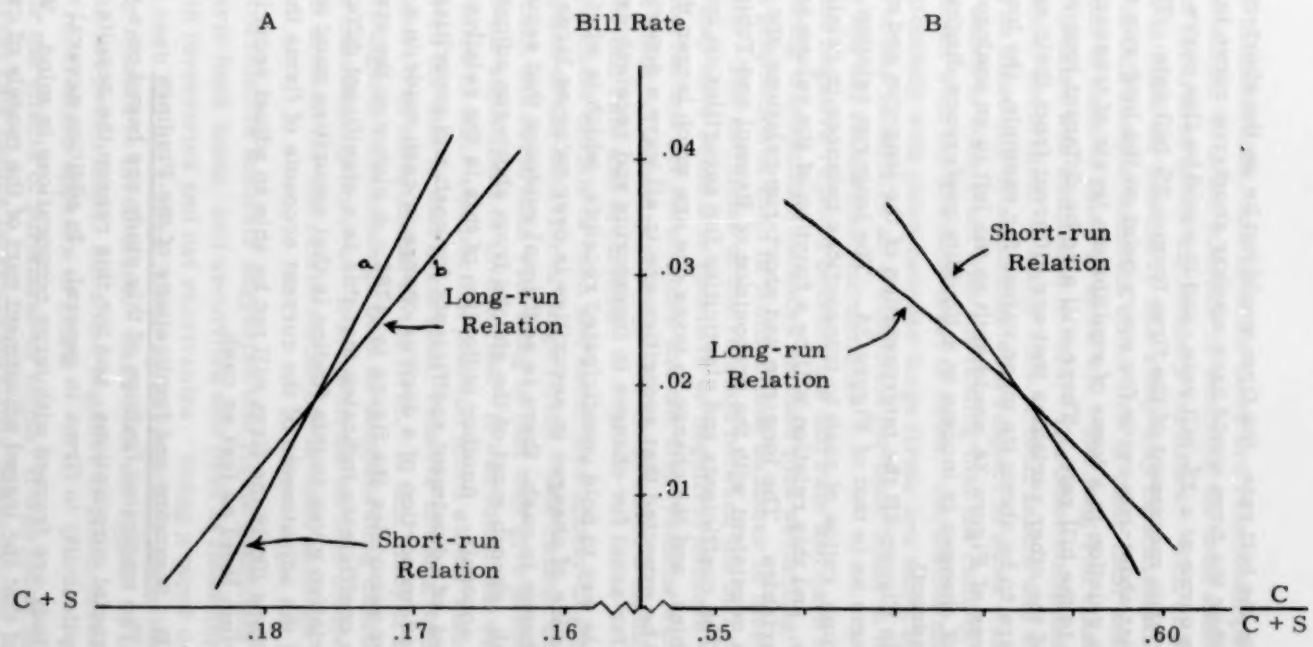
1. The equality of the two estimates of the percentage decrease in cash could be estimated if the variances of these estimates were homogeneous. When the variances of the two estimates were compared using Bartlett's test, the Chi Squared value was over 100 indicating that the variances were not homogeneous. It may be noted that the present long-run estimate is within two standard deviations of Kisselgoff's estimate of the percentage decrease in cash due to a 1% rise in the interest rate. However, because the number of observations is much larger in this study the standard error of the present estimate would get most of the weight in tests of equality of the two estimates.

2. See footnote 2, p. 153 above for a discussion of Kisselgoff's interest estimates. A source of bias for the present estimates, the use of quantities rather than returns on inventories and net receivables, is discussed above, p. 145.

3. See above, footnote 1, p. 120–121.

Figure 3

Estimated Relation of Transactions Balances and the Proportion of Cash in Transactions Balances to the Bill Rate



change in the bill rate, the firm would not be on the short-run curve of Figure 3A. Rather the firm would have another short-run curve intersecting the long-run curve at a 3% bill rate, and it would be this curve that would show the short-run movement of the firm from a 3% bill rate. There will thus be a different short-run curve for every point on the long-run relation. The long-run relation is a locus of equilibrium levels of transactions balances in relation to the bill rate. There will also be different long-run relations for values of the other variables that are different from their mean values. If sales were to be above its mean value, for example, the long-run and short-run curves of Figure 3A would shift to the left in an analogous manner to the effects of changes in income in a liquidity preference diagram (in the northwest quadrant).

In Figure 3B the interpretation of the long-run and short-run relations is the same as to that of Figure 3A. The long-run relation is a locus of equilibrium ratios of cash to transactions balances in relation to the bill rate, and this relation will be a function of the values assumed for the other variables. The long-run and short-run relations are negatively sloped which is consistent with the proposition of Baumol and Tobin.

The coefficients and elasticities for securities on inventories, net receivables, and the change in taxes due are much larger than those for cash. It would be expected that securities are in all ways a superior asset to hold as a buffer asset for changes in inventories and receivables, and a more profitable way to hold unanticipated receipts, which is what was found. As the variance of changes in securities is over twice as large as the variance of the change in cash, there is additional evidence that securities, rather than cash, absorb most of the shocks from short-run adjustments in the current accounts. Another indication of this is the relative magnitudes of the speed of adjustment coefficients for cash and securities. For cash the estimated proportion of a desired change in cash made in a year is .6312, while for securities the figure is .4782. A glance at the standard errors of these coefficients indicates that this is a significant difference. The interpretation given to this finding is that securities must absorb much more of the adjustments of the current accounts of firms than cash and it is therefore likely the firm will not be able to adjust securities to their equilibrium level as fast as cash.

B. Summary and Implications of the Findings

The empirical findings of this study are based on a sample of large non-financial corporations, and for this reason the results are unlikely to have applicability to firms in general. In addition several of the hypotheses of the study are framed with large corporations in mind. While it might be expected that the lagged adjustment part of the models of cash and securities would describe a large class of firms, it is doubtful whether the relationship between securities and inventories suggested for large corporations would have much applicability to smaller firms.

A serious limitation on the analysis is that it is partial. Because

the models are not simultaneous, and because there are relationships among the independent variables, it is not possible to describe the total short-run impact on cash and securities of the explanatory variables. With at least these reservations in mind the results of this study may be summarized.

The firms in this sample maintain about 10% of their total assets as cash, while the ratio of cash to sales is about 6%. The long-run elasticity of cash with respect to sales was estimated at .885. The other component of transactions balances, securities, is about 7.3% of total assets and 4.2% of sales. The long-run elasticity of securities with respect to sales was estimated to be 1.31, which is consistent with the finding that the ratio of securities to sales increases with the absolute level of sales. Securities may rise more than proportionately with sales because large firms are more likely to find it economical to use securities as a substitute for bank loans in the finance of inventory and net receivable fluctuations.

The expected negative relation between the bill rate and cash holdings was found. Measuring transactions balances as cash plus securities, it was also found that the proportion of cash in transactions balances is inversely related to the bill rate. The estimates of the relation between the bill rate and cash holdings may have some bias (positive or negative) because the quantities of inventories and net receivables and not their returns were used as variables. The model of security holdings did not produce a significant relation between the bill rate and securities.

Lagged adjustment models of cash and securities appear to give a reasonable description of the behavior of large firms. One reason that firms may adjust cash and securities with a lag is that they give priority in the short-run to the adjustment of inventories. It was found that securities tend to be adjusted to their equilibrium position more slowly than cash. One reason suggested for this is that securities, relative to cash, tend to bear the brunt of short-run adjustments of the firm. In this study it has not been possible to distinguish cash holdings held as compensating balances in banks where firms customarily borrow. If large firms do hold compensating balances this might account for the relative stability of cash holdings that has been found.

Canonical correlation analysis suggested that firms substitute between the use of short-term bank loans, and securities and cash in financing temporary increases in inventories and net receivables. Using a more complete model to describe cash and securities it was again found that securities and cash are negatively related to inventories. Cash does not appear to be significantly related to net receivables, but some negative relation between securities and net receivables appears to exist.

Allowing for the effects of purchases by firms, it was found that increases in the interest rate induce large corporations to expand the credit they take in the form of accounts payable. However, the relation between the interest rate and net trade credit taken by the customers of these firms, given the effects of time and sales, is even more striking. The estimates suggest that net receivables as a percentage of total assets increase by about 1% in

response to a rise in the bill rate of 1%, say from 2% to 3%. In addition the net receivables of large corporations do increase with the sales of firms which means that during an expansion of business activity there will be a sizeable amount of credit offered by large firms to their customers. During a business expansion interest rates and inventories will also be rising. In response to the higher interest rates large firms reduce their cash needs significantly, and some part of the increase in net receivables and inventories is financed by running down security and cash holdings.

These results suggest that during an inflationary period large corporations are in a position to finance some of their credit needs internally, and to extend credit to smaller businesses and consumers.[1] As these large corporations do use short-term bank loans to finance a substantial part of their increase in net receivables and inventories during an expansion, there is clearly scope for restraining their activity. However, if, because of earnings, low risk, or other reasons, large firms are more likely to be accommodated by the banking system in periods of tight money than smaller borrowers, the scope for restraining their activity is substantially reduced. Assuming it is true that large corporations are somewhat removed from the impact of monetary policy because they maintain internal sources of finance and are accommodated readily by the banking system, and if it is desired to influence their behavior, what can be done about it? In attempting to answer this question the results of this study are interpreted more liberally than is probably justified.

The problem could be attacked from the standpoint of the availability of short-term credit to large firms, but this approach appears to involve discriminatory controls on bank lending. In addition, firms, in the face of tight money, are able to expand their activity by the sales of securities to the extent that they have adequate long-term credit. During a business expansion it would probably require a very high bill rate to induce firms to retain their security holdings instead of increasing net receivables and inventories. Another approach to the problem is to view large corporations for present purposes as a type of financial intermediary. As it may be difficult to control the source of funds of large corporations, attention is focused on the control of their lending.

An alternative method of restraining large corporations is to curtail

1. Alan Meltzer, on the basis of aggregative data for firms of various sizes, draws inferences directly related to the above discussion. Meltzer suggests that, because large firms extend sizeable amounts of trade credit to small firms in periods of tight money, small firms are able to obtain credit from banks indirectly, even though credit rationing directly favors large firms. Unfortunately, Meltzer's study was read after the present paper was written so that it has not been possible to compare his more detailed findings with those of Section 3 of this study. See "Mercantile Credit, Monetary Policy and the Size of Firms," Review of Economics and Statistics, November, 1960.

the trade credit they offer their customers. If the terms on accounts receivable were subject to control, such as specification of the minimum terms that could be offered, the monetary impact of large corporations might be significantly reduced. If monetary restraint on some portion of the customers of large firms is effective, then the effect of tightening terms on receivables will be to reduce the sales of all firms including large corporations. This will in turn reduce the demand for inventories of large corporations, and result in a lower level of activity. Symmetrically, if the monetary authorities reduced the terms in recessions, it might prove a stimulant to sales and inventory investment. These comments are intended to be provocative of further thought and research. Much more study is needed to determine whether large corporations are fairly insulated from the effects of monetary policy, and whether the trade credit offered by large firms allows their customers to make purchases they would not otherwise have been able to finance. Also some judgment must be made as to the desirability of the use of selective controls of the type suggested.

Appendix 1: Analysis of Firm Intercepts

The firm intercepts for equations (4-11) and (4-12) for cash and securities have been estimated. Examination of these equations reveals that the estimated intercepts are a product of two coefficients, $\beta_3\alpha_{60}$ for cash and $\beta_4\alpha_{70}$ for securities. To obtain an estimate of α_{60} and α_{70}, which are the intercepts in the equations for desired cash (4-1) and desired securities (4-2), the calculated intercepts for each firm are divided by the estimate of β_3 for cash and β_4 for securities. The purpose of estimating these coefficients was to see if they exhibited any pattern by industry or asset size.

In Table A-1 the following statistics are presented. In column (1) the mean of average cash holdings by industry is presented, in column (2) the mean of α''_{60} by industry, in column (3) the mean of average cash holdings by industry divided by the standard deviation, and in column (4) the mean of α''_{60} by industry divided by its standard deviation. In columns (5) to (8) the same statistics for securities are presented. These same statistics are presented for the firms classified by asset size. The mean intercepts for cash, and even more so for securities, do tend to exhibit a pattern similar to that of the means of average cash and security holdings both by asset size and industry. The coefficients of variation of the intercepts for cash and securities are of similar size for the various groups of firms, while the coefficients of variation for the means of cash are generally much larger than the coefficients of variation for the intercepts for cash.

If for any of the groups of firms examined the intercepts clustered about some mean value it would suggest that only an intercept for the whole group of firms need be estimated, rather than an intercept for each firm. The results in Table A-1 are not very suggestive on this count. The possible exceptions are industry groups 3, 4, and 5, whose intercepts for both cash and securities display much less variation than other groups of firms.

TABLE A-1

Means and Coefficients of Variation (V) of Cash and Security Holdings and Firm Intercepts by Asset Size and Industry

	Mean of Cash (1)	Mean of α''_{60} (2)	V for Cash (3)	V for α''_{60} (4)	Mean of Securities (5)	Mean of α''_{70} (6)	V for Securities (7)	V for α''_{70} (8)
Asset Size Classes								
1. $10 to 40 million	.1222	.0432	2.23	.97	.0548	.0324	.79	.33
2. $40 to 65 million	.1080	.0089	2.65	.10	.0671	.0234	1.17	.18
3. $65 to 100 million	.1414	.0921	2.03	1.51	.0528	.0658	.71	.75
4. $100 to 150 million	.1071	.0560	2.65	.86	.0581	.0819	1.08	1.06
5. $150 to 200 million	.0849	.0388	3.64	1.37	.1101	.1315	.89	.90
6. $200 to 400 million	.1027	.0424	2.59	.88	.0654	.0546	1.08	.55
7. $400 to 700 million	.1040	.0518	1.75	.65	.0769	.0916	1.32	.87
8. Over $700 million	.0742	.0339	2.13	1.25	.0979	.0962	1.37	1.13
Industry								
1. Food and Tobacco	.1114	.0077	1.68	.06	.0473	.0102	.68	.07
2. Petroleum and Rubber	.0729	.0366	2.40	1.20	.0591	.0483	1.37	.88
3. Steel and Non-Ferrous	.0821	.0445	2.68	1.21	.0940	.1066	1.24	1.20
4. Chemicals	.1048	.0717	3.25	1.80	.0993	.1160	1.55	1.64
5. Machinery	.1020	.0680	2.46	1.89	.0843	.1262	1.00	1.25
6. Transportation	.1131	.0637	3.48	2.40	.0920	.1155	.95	1.39
7. Retail Trade	.1326	.0275	2.04	.41	.0522	.0045	.82	.05
Total	.1042	.0446	2.32	.71	.0734	.0710	.98	.66

Appendix 2: Analysis of Residuals by Year, Asset Size and Industry of the Firms, and Size of the Independent Variables

In Section 4 it was mentioned that the analysis of the residuals is based on a slightly different form of the regression equations than was presented in that section. The appendix will begin with a discussion of this point. With no loss of generality, and some saving of space, we will consider for illustrative purposes that the change in cash is related to one independent variable, X, and that separate intercepts are specified for each firm so we have

$$(A-1)\quad \Delta C_{it} = \alpha_{1i} + \alpha_2 \overline{X}_{it} + u_{it}$$

where, as in the text, t runs over 9 years and i over 165 firms. When least squares estimating procedure is used,

$$(A-2)\quad \alpha_{1i} = \overline{\Delta C}_i - \alpha_2 \overline{X}_i .$$

Substituting (A-2) for α_{1i} in (A-1) we have

$$(A-3)\quad (\Delta C_{it} - \overline{\Delta C}_i) = \alpha_2 (X_{it} - \overline{X}_i) + u_{it} .$$

Thus (A-3) is an alternative way of formulating an equation specifying separate intercepts.

The reason given in Section 4 for specifying separate intercepts is that it is frequently difficult to explain the differences in the average value of stock variables among firms. However, means of changes in stocks (or flows) like $\overline{\Delta C}$ are not apt to exhibit many inexplicable differences among firms, at least over a short period. Thus instead of (A-3) one might write

$$(A-4)\quad \Delta C_{it} = \alpha_2 (X_{it} - \overline{X}_i) + \alpha_3 + v_{it},$$

where the v's are a new error term, and a common intercept, α_3, is specified. If we consider X to be a stock or flow variable whose average value among firms is not easily explained, then it will be useful to use this variable as a deviation from the mean of each firm, especially since it costs no degrees of freedom to do this.

Interestingly the least square estimates of α_2 are the same in (A-4) as in (A-3), and therefore as in (A-1). In (A-3) and (A-4) the variance of $(X_{it} - \overline{X}_i)$ will be identical. Since $\Sigma\Sigma(\Delta C_{it} - \overline{\Delta C}_i)(X_{it} - \overline{X}_i) = \Sigma\Sigma \Delta C_{it} (X_{it} - \overline{X}_i)$, the covariance is the same in (A-3) and (A-4), and so the two estimates of α_2 will be the same. For any number of X's the matrix of raw moments for form (A-4) will be identical with the form (A-3) except for the row of sums.[1] Instead of the number of observations and zeros in the row of sums as in (A-3), the row of sums for form (A-4) will consist of the number of observations, zeros for all the independent variables, and the last element will be the sum of ΔC_{it}. So the slope estimates on the independent variables will be the same in (A-4) as in (A-3), and the inverse of

1. And except for the measure of the variance of the change in cash in the last row.

the matrix of raw moments will be the same. The estimate of the common intercepts, α_3, will be precisely $\overline{\Delta C}$. Finally the difference in residual variance between form (A-3) and (A-4) will be $t\Sigma(\overline{\Delta C}_i - \overline{\Delta C})^2$. Since for changes in stocks or flows this magnitude is not apt to be large, it points to the advantage of form (A-4) over (A-3). However, when a firm stock or flow is the dependent variable, the situation may be different. [1]

Another advantage of form (A-4) over (A-3) is saving degrees of freedom. In the present study, going from form (A-4) to (A-3) uses 164 additional degrees of freedom in estimating the mean of ΔC for each firm. The estimates of the cash and security equations of Section 4 when written in the form of (A-4) are

$$(A-5)'' \quad \Delta C_{it} = \underset{(.0000)}{.0020} + \underset{(.0029)}{.0330}\,(T_{it}-\overline{T}_i) - \underset{(.0118)}{.3848}\,(B_t-\overline{B}) - \underset{(.0118)}{.0391}\,(I_{it}-\overline{I}_i)$$

$$- \underset{(.0151)}{.0098}\,(R_{it}-\overline{R}_i-D_{it}+\overline{D}_i) + \underset{(.0235)}{.6312}\,[(\overline{C}_{t-1})_i - C_{it-1}] + .1207\,(\Delta N_{it}-\Delta\overline{N}_i) + w_{3it},$$

$$(A-6)'' \quad \Delta S_{it} = \underset{(.0000)}{.0029} + \underset{(.0041)}{.0263}\,(T_{it}-\overline{T}_i) + \underset{(.1942)}{.0008}\,(B_t-\overline{B}) - \underset{(.0174)}{.1321}\,(I_{it}-\overline{I}_i)$$

$$- \underset{(.0222)}{.0260}\,(R_{it}-\overline{R}_i-D_{it}+\overline{D}_i) + \underset{(.0225)}{.4782}\,[(\overline{S}_{t-1})_i - S_{it-1})] + \underset{(.0301)}{.3237}\,(\Delta N_{it}-\overline{\Delta N}_i) + w_{4it}.$$

As mentioned, the coefficients of the independent variables are identical with those of equations B of Table 4, but now there are common intercepts of $\overline{\Delta C}$ and $\overline{\Delta S}$ respectively. The coefficients above have lower standard errors because the sum of residuals squared when corrected for degrees of freedom is less in (A-5) and (A-6) than in equations B of Table 4. $R^2 = .371$ and $F = 145.82$ for cash in (A-5) and $R^2 = .352$ and $F = 126.58$ for securities in (A-6). As the same degrees of freedom are used in estimating (A-5) and (A-6) as are used in estimating equations A of Table 4, it is clear that (A-5) and (A-6) explain significantly more of the variance of the change in cash and the change in securities. The addition of intercepts for each firm to (A-5) and (A-6) does not add significantly to the explained variance of changes in cash and securities.

Having established why w_{3it} and w_{4it} of (A-5) and (A-6) are the residuals that are to be examined, some general comments about the character of these error terms are in order. From the form of (A-4) it should be clear that the sum of residuals for each firm for cash is $t(\overline{\Delta C}_i - \overline{\Delta C})$, and for securities, $t(\overline{\Delta S}_i - \overline{\Delta S})$. As the sum of residuals for each firm are unrelated to the independent variables, which sum to zero

---

1. See p.138, Section 3, above where two stocks, payables and net receivables, were used as dependent variables.

for each firm, they are not of much interest. If the sums of residuals for each firm are of little interest, then the same is true for groups of firms by asset size and industry. However, the sum of residuals by year will be of interest and this is discussed below.

For this reason the analysis deals principally with the residual variance of subgroups of observations. Suppose for cash the first ten firms of the sample are a subgroup of interest. The initial variance of ΔC for this group is $\sum_{1}^{10}\sum_{47}^{56}(\Delta C_{it} - \overline{\Delta C})^2$, and subtracting from this $\sum_{1}^{10}\sum_{47}^{56} w_{3it}^2$ we get a quantity which we shall call the "explained variance" for a subgroup. We then make the usual F test. The explained variance is divided by 6, for the 6 independent parameters estimated, to give the numerator of the F ratio. The denominator is the sum of residuals squared for the subgroup divided by the number of observations less 7, the number of parameters estimated. Thus, for the first ten firms of the sample,

$$F = \frac{\left[\sum_{1}^{10}\sum_{47}^{56}(\Delta C_{it} - \overline{\Delta C})^2 - \sum_{1}^{10}\sum_{47}^{56} w_{3it}^2\right] \div 6}{\left[\sum_{1}^{10}\sum_{47}^{56} w_{3it}^2\right] \div (90-7)}$$

The principal peculiarity of our concept of explained variance for a subgroup is that it could be negative. In this case, of course, no test is needed, and the hypothesis that the explained variance for the subgroup is zero (or less) is accepted. If we summed the numerator of the F ratios for an exhaustive set of subgroups, it would equal the numerator of the F ratio for the regression equation for all observations. However, the sum of the denominator over all subgroups must necessarily be larger than the denominator for the F ratio of the regression equation for all observations.[1] This means that the null hypothesis that the explained variance of a subgroup is zero is more likely to be accepted for a subgroup, regardless of the difference in observations, than the same hypothesis for all observations. Since we want to be able to reject the null hypothesis this characteristic of the test is (rightly?) not in our favor.

In Table A-2 firms are classified into 8 asset size groups and 7 industry groups. The explained variance as defined above and the residual

1. This characteristic of the test makes it very difficult to reject the null hypothesis for small subgroups. Several of the industrial groupings of the present sample were four firms or less, and these have been put with other industrial groups for this reason. Thus, autos are included with transportation equipment, tobacco with food, rubber with petroleum, and non-ferrous metals with steel.

TABLE A-2

Analysis of Residual Variance by Asset Size and Industry

	Sum of Squared Residuals (1)	Cash Explained Variance (2)	F* (3)	Sum of Squared Residuals (4)	Securities Explained Variance (5)	F* (6)	Number of Firms (7)
A. Asset Size Groups							
1. $10 to 40 million	.20615	.07627	10.77	.18288	.19009	30.45	20
2. $40 to 65 million	.16822	.14530	23.73	.28366	.15121	14.73	19
3. $65 to 100 million	.21210	.17961	20.78	.19245	.06469	8.35	17
4. $100 to 150 million	.10641	.08139	19.94	.24623	.07598	8.06	18
5. $150 to 200 million	.10245	.07798	18.52	.38345	.21051	13.54	17
6. $200 to 400 million	.11928	.03307	10.20	.36922	.19304	19.14	25
7. $400 to 700 million	.16023	.04254	9.91	.32298	.13472	14.67	24
8. Over $700 million	.06148	.03612	21.49	.48688	.31354	29.18	25
B. Industry							
1. Food and Tobacco	.10721	.05081	17.27	.17590	.05093	10.74	25
2. Petroleum and Rubber	.05867	.01677	8.73	.12151	.07547	19.34	21
3. Chemicals	.16168	.12747	24.13	.49781	.24401	15.00	21
4. Steel and Non-Ferrous	.12180	.04922	11.71	.27349	.22391	23.91	20
5. Machinery	.34118	.25299	36.97	.75302	.35925	23.75	34
6. Transportation Equipment	.11570	.08158	14.15	.40303	.29538	14.69	14
7. Retail Trade	.23008	.09380	17.96	.23632	.08716	16.31	30

* Critical values of F range from 2.90 (2.14) for industry group 5 with 34 firms (306 observations) to 2.99 (2.19) for industry group 6 with 14 firms (126 observations), at the 1% (5%) level.

variance for these groups is presented in columns (1) and (2) for cash and (4) and (5) for securities. The above F statistics are given in column (3) for cash and (6) for securities. For all subgroups the explained variance is significantly different from zero at the 1% level. It should be made explicit that these tests do not tell us whether it would have been "better" to run a separate regression for each subgroup.

We turn now to analysis of residuals by year. In Table A-3 the sums of residuals for cash and securities and the sums of the dependent variables, ΔC and ΔS, are presented by year in columns (1) and (2). In columns (3) and (4) the residual variance and the explained variance are presented, and in column (5) the F ratios are given. It may be noted that the sums of the estimated residuals by year will be the same for either equation form (A-3) or (A-4).

Do the means of residuals differ by year? This question may be answered if it is assumed that the variances of residuals are the same in each year, an assumption of the present regression model. The critical value of F for this test for 8 degrees of freedom in the numerator (8 independent estimates of means) and (1485-7-9) degrees of freedom in the denominator is 2.53 (1.95) at the 1% (5%) level of significance. The F ratio for cash is 1.57, and the null hypothesis that the means of the residuals for cash are equal, and equal to zero, is accepted. The F ratio for securities is 3.58 and the null hypothesis is rejected for securities. The null hypothesis that the explained variance for each year is zero is rejected in all years except 1948 for securities and 1956 for cash.

The results that the means are different from zero for securities has something to do with time. The correlation of time with the residuals of each firm is not significant at the 1% level, and the correlation with the sums of residuals is not significant at either the 5% or 1% levels. However, the correlation is enough to substantially change the patterns of the sums of residuals. If we estimate

$$w_{4it} = a(t-\bar{t}) + w'_{4it},$$

the estimate of "a" is .0009, R^2 = .003, and the F ratio is 4.83. The critical value of F is 3.84 at the 5% level and 6.66 at the 1% level. If time is used to explain the residuals the sum in 1948 is -1.1, and in 1955 is .67, which is an improvement, but the sum in 1956 becomes -1.38, and in 1949, 1.14. For the test of the homogeneity of means the F value for securities with time as a variable is 2.97, and the hypothesis that the means are equal in each year would be rejected at the 5% or 1% level. However, whereas the null hypothesis that the explained variance for securities was zero at the 5% or 1% level in 1948 was accepted when time was not a variable, with time as a variable the F ratio would be 2.97, which is greater than the critical value at the 1% level, i.e., 2.72, with 7 and 158 degrees of freedom. These results suggest, we believe, that the explanatory value of time for securities has something to do with 1948.

In 1948 the equation predicted an increase in securities for 152 of

TABLE A-3

Analysis of Residual Variance by Year*

Year	Sum of Dependent Variable (1)	Sum of Residuals (2)	Sum of Residuals Squared (3)	Explained Variance (4)	F Ratio** (5)
1948					
Cash	-.217669	-.644198	.12688	.10213	22.30
Securities	.973297	-1.700487	.27179	.02192	2.14***
1949					
Cash	.532040	.585148	.10799	.03699	9.06
Securities	2.073883	.691080	.22374	.08990	10.62
1950					
Cash	.564818	-.055149	.11589	.03851	8.78
Securities	1.587197	-.617922	.18500	.07036	10.11
1951					
Cash	.785820	-.237122	.16306	.13436	21.73
Securities	.584626	-.706728	.35175	.21132	15.93
1952					
Cash	.108225	.321343	.11727	.08989	20.23
Securities	-.813031	.884847	.29845	.38934	34.50
1953					
Cash	.913694	-.122216	.16000	.07639	12.60
Securities	.715271	.214770	.26881	.11043	10.89
1954					
Cash	.197350	-.356185	.11160	.09072	21.59
Securities	-.323810	.898786	.22307	.09014	10.73
1955					
Cash	-.579722	-.608756	.10181	.09117	26.50
Securities	1.689740	1.118945	.31649	.12693	10.63
1956					
Cash	.676442	.404765	.13186	.00741	1.48***
Securities	-2.233863	-.783291	.31605	.22596	18.92

* Figures in columns (2) through (4) need to be divided by 165 to obtain average values for any year.

** Critical values of F for each year are 2.92 (2.16) at the 1% (5%) level.

*** Null hypothesis is accepted at 5% or 1% level of significance.

TABLE A-4

Analysis of Residual Variance of Firms Classified by Size of Sales, Inventories, Cash, and Securities (Range of variable for each classification given with class numbers)*

Variable of Classification	Index of Residual Variance: Cash (1)	Securities (2)
A. Sales		
1. .54- .77	89.8	70.5
2. .77- .91	72.1	115.5
3. .91-1.02	65.4	186.9
4. 1.02-1.21	79.3	127.2
5. 1.21-1.36	95.0	79.5
6. 1.36-1.43	82.8	124.7
7. 1.43-1.69	114.7	123.7
8. 1.69-2.16	164.5	63.6
9. 2.16-2.77	133.9	83.1
10. 2.77-6.11	100.0	30.2
B. Inventories		
1. .046-.124	56.8	82.9
2. .124-.160	82.5	97.9
3. .160-.194	127.8	119.9
4. .194-.246	72.8	128.0
5. .246-.276	103.2	143.8
6. .276-.305	98.8	97.9
7. .305-.341	84.1	92.3
8. .341-.380	146.5	69.8
9. .380-.456	157.8	57.4
10. .456-.864	68.5	57.9
C. Lagged Cash		
1. .008-.050	16.1	
2. .050-.065	38.6	
3. .065-.073	36.1	
4. .073-.079	40.8	
5. .079-.089	56.1	
6. .089-.102	120.8	
7. .102-.124	114.8	
8. .124-.138	167.1	
9. .138-.175	200.3	
10. .175-.317	194.2	
D. Lagged Securities		
1. .000-.001		13.0
2. .001-.007		26.7
3. .007-.022		30.2
4. .022-.030		54.8
5. .030-.045		87.4
6. .045-.079		110.0
7. .079-.099		118.1
8. .099-.126		174.4
9. .126-.158		182.5
10. .158-.482		186.8

* Sixteen firms are included in each of the first five classes for each variable, and 17 firms in the last five classes. The average of the sum of squared residuals over all observations was used to form the index numbers. Thus the index numbers for cash for all variables have the same base and similarly for securities.

the 165 firms. This preponderance of predicted increases may be partly due to the trend in the values of variables. However, inventory holdings for the firms in 1948 were not high relative to sales, so this factor led to more predicted increases in 1948 relative to other years of business expansion, and the fact that the sum of changes in securities is positive in 1948 also suggests that an equation describing security holdings might well have a large number of predicted increases in that year. However, actual increases in security holdings were made by only 70 firms. The problem with 1948, we suggest, is that a substantial number of the firms were still getting rid of their accumulation of securities from the war. This adjustment, though related to the lagged value of securities, would not necessarily be related to the values of the other variables in the model, and may be the reason for the relative poor performance of the securities equation in 1948.

The performance of the cash equation in 1956 is somewhat mystifying. An examination of signs revealed that the predicted change in cash holdings in 1956 was of the same sign as the actual change for 105 of the firms. Neither the sums of the changes in cash or the sums of residuals are suggestive of any particular bias in that year. The values of the independent variables for that year are also not suggestive of why the explanatory value of the regression equation is low for 1956. One reason for the poor performance of the cash equation for 1956, which is not helpful in indicating the manner in which the estimates err, is that the initial variance of changes in cash is lowest in that year.

The error terms for cash and securities have also been related to the size of some of the independent variables, and in particular, sales, inventories, lagged cash, and lagged securities. As the size of these independent variables for each firm will tend to rise through time because they are divided by average total assets for the period, the average values of these variables for each firm have been used to rank firms. That is, firms have been ranked by the average size of the independent variables, and not observations. The firms have been divided into ten groups by the size of the variables, and the average residual variance has been converted into an index number for each group. These index numbers are presented in Table A-4.

It is quite obvious that there is substantial heteroskedasticity for cash with respect to lagged values of cash, and for securities with respect to lagged securities. For sales, it is not easy to tell whether the pattern of residual variance suggests a non-linear relationship between cash and sales, or that there is simply heteroskedasticity. For inventories the pattern of residual variance is quite mixed for cash. For securities the residual variance for both sales and inventories suggests that there may be some type of non-linear relationship which would describe the data better than the simple linear forms used here.

Summary: This study examines the characteristics of speculation in the exchange markets under flexible exchange rate systems. The experiences of five European countries—Great Britain, France, Belgium, the Netherlands, and Switzerland—with flexible rate systems from 1919 to 1926 are evaluated to see whether reliance on market forces for the determination of the rate results in smooth, orderly rate movements which conform with the changes in the equilibrium rate.

Considerable attention is given to changes in the differential between the spot exchange rate and the forward exchange rate as an indication of how speculative anticipations change through time, and whether transactions of forward speculators amplify or dampen the movements in the spot rate. The data examined do not show that forward speculators dampened rate movements or that they gave support to currencies which were temporarily weak in the exchange market. They suggest that forward speculators amplify movements in the spot rate as the level of the spot rate changes.

Speculators in the forward market showed little of the flexibility which would be required if they were charged with keeping rate movements small and orderly. Instead, their anticipations were dominated for a long period by the prewar parity relationships rather than by current economic events. In Britain, speculators moved the actual rate back toward its prewar parity, causing sterling to become greatly overvalued. In France, the turnabout in speculative expectations after a belated realization that franc would not be restored to its prewar parity resulted in a depreciation of the franc of fifty percent in twenty weeks. The subsequent interaction of a depreciating currency and rising prices was so quick that self-limiting, automatic checks to further decline in the exchange rate proved ineffective. In Belgium, the successful effort in achieving internal stability was thwarted because the speculative belief that the French franc and Belgian franc should exchange on a one-for-one basis forced the Belgian franc to decline with the French franc. In all these cases, speculative transactions caused the actual rate to deviate significantly from the equilibrium rate, with unfortunate consequences, so that the extra degree of freedom that a flexible rate system offered was greatly compromised.

This essay was submitted to Yale University in 1961 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy, to be awarded in 1962. The author is currently a staff economist at the Committee for Economic Development.

Dissertation Committee: Professors Robert Triffin (Chairman), Henry C. Wallich, and Gustav Ranis.

SPECULATION IN THE FOREIGN EXCHANGES: THE EUROPEAN EXPERIENCE, 1919–1926

Robert Z. Aliber

1. Introduction

One of the points of controversy in the arguments for and against a system of flexible exchange rates is whether private speculative transactions are likely to be stabilizing or destabilizing.[1] The case for flexible rates is attractive, for if the authorities do not have to maintain a particular exchange rate they may find it easier to adopt the monetary and fiscal policies best suited to secure their domestic economic objectives. But if the removal of the official support for a currency at a particular level subjects it to sizeable fluctuations in the exchange markets, there may be an undesirable reduction in the level of foreign trade and investment. Moreover, if speculative transactions have a significant influence on the level of the exchange rate, they may have undesirable impacts on price and income movements in the countries whose exchange rates are so affected. If, however, speculative transactions prove stabilizing, both of these undesirable effects are likely to be avoided.

The basis for believing that speculation is likely to be destabilizing was initially an empirical one based on interpretations by various economists of the experience of various European countries with flexible exchange rate systems after World War I.[2] These authors gave two meanings to the term

---

1. A flexible, floating, or free exchange rate system relies primarily on private transactions for the determination of the foreign exchange rate. Official transactions undertaken to make payments abroad are not intended to affect the level of the rate.

Speculative foreign exchange transactions are undertaken in anticipation of a profit from an expected change in the exchange rate. Such transactions may be undertaken by traders who alter the timing of the exchange transactions associated with their commercial transactions, or they may be undertaken by others whose only reason for engaging in foreign exchange operations is the hope of a speculative profit.

Stabilizing transactions reduce the extent of the deviation of the actual rate from an equilibrium rate; destabilizing transactions, conversely, increase the extent of the deviation.

2. The classic source is Ragnar Nurkse. "If there is anything that the inter-war experience has clearly demonstrated, it is that paper currency exchanges cannot be left free to fluctuate from day to day under the influence of market supply and demand . . . If currencies are left free to fluctuate 'speculation' in the widest sense is likely to play havoc with the exchange rates . . . Any considerable or continuous movement of the exchange rate is liable to generate anticipations of a further movement in the same direction, thus giving rise to speculative capital transfers of a disequilibrating kind tending greatly to accentuate any change that may be

"destabilizing speculation." One was that speculative transactions amplified the rate fluctuations, causing the actual rate movement to overtrack the movement which otherwise would have prevailed; the second, that speculative transactions caused some currencies to become excessively depreciated, and others to become excessively appreciated, to an extent much greater than was warranted by the domestic economic situation.

This empirical argument has been supplemented by a simple theoretical model which conforms to the speculator's maxim, "Run your gains, cut your losses." This model was designed to show that speculative transactions can be destabilizing and profitable at the same time.[1] In this model speculators purchase an appreciating currency as long as it continues to appreciate. Eventually, despite their purchases, it reaches a turning point and begins to depreciate, and then they sell it. But their continued purchases of this currency as it appreciated causes this turning point to be higher than it would have been otherwise. Thus they have caused the rate movement to overtrack.

Those who believe that speculation is likely to be stabilizing also support their case on both empirical and theoretical grounds. It is sometimes asserted that exchange speculation has facilitated adjustments in the Canadian dollar in the last decade, and has thus been stabilizing.[2] Moreover, it is maintained that speculation should be considered stabilizing in some cases

---

required for the balancing of normal transactions. Moreover, the normal transactions also may come to be affected by speculative anticipations." International Currency Experience (Princeton: League of Nations, 1944), pp. 137-138, p. 210.

Others include Eleanor Lansing Dulles, The French Franc, 1914-1928 (New York: Macmillan & Co., 1929); J. van Walre de Bordes, The Austrian Crown (London: P. S. King & Son, Ltd., 1924); Henry L Shepherd, The Monetary Experience of Belgium, 1914-1936 (Princeton: Princeton University Press, 1936); Frank D Graham, Exchange, Prices, and Production in Hyper-Inflation Germany, 1920-1923 (Princeton: Princeton University Press, 1930); Constantino Bresciani-Turroni, The Economics of Inflation (London: George Allen & Unwin Ltd., 1931); James Harvey Rogers, The Process of Inflation in France, 1914-1927 (New York: Columbia University Press, 1929); and Commission on Gold and Silver Inquiry, United States Senate, European Currency and Finance, Washington, 1925.

1. W. J. Baumol, "Speculation and Economic Stability," Review of Economics and Statistics, August 1957, pp. 263-271.

2. H. C. Eastman and S. Stykolt, "Exchange Stabilization in Canada," Canadian Journal of Economics and Political Science, May 1956, pp. 221-233, and "Exchange Stabilization Further Considered," Ibid., August 1957; H. C. Eastman, "Aspects of Speculation in the Canadian Market for Foreign Exchange," Ibid., August 1958, pp. 355-372; R. R. Rhomberg, "Canada's Foreign Exchange Market: A Quarterly Model," Staff Papers, April 1960, pp. 439-456.

in the interwar period because speculators were only anticipating the weakness of the currencies resulting from non-speculative factors.[1]

The theoretical argument is based on the premise that, if speculation is profitable, then it very likely is stabilizing.[2] Where the equilibrium rate is constant, it is argued that profitable speculation would dampen rate fluctuations arising from non-speculative factors, and where the equilibrium rate changes through time, that profitable speculation would quicken the adjustment to the new equilibrium rate.[3]

Thus the controversy has been stalemated at both the empirical and the theoretical levels. At the empirical level, the failure to join the issue has resulted from the tendency to base generalizations on different experiences with flexible rate systems, as well as a tendency to draw different conclusions from the same experience. At the theoretical level, the failure to join the issue reflects differences in the assumptions about speculative strategies, in particular about which strategy speculators are likely to adopt in their search for profits.

Both the proponents and critics of flexible rate systems generally develop models for situations where the equilibrium rate is unchanged. This is somewhat paradoxical, for if the equilibrium rate remains unchanged the need for a flexible rate becomes academic. The proponents of flexible rate systems acknowledge this criticism, defend it on the ground of pedagogic convenience, and maintain that the basic argument remains unaffected as long as speculators can be reasonably certain about the future level of the equilibrium rate.[4]

---

1. Milton Friedman has commented, "Nurkse concludes from interwar experience that speculation can be expected in general to be destabilizing . . . Nurkse examines only one period in anything approaching the required detail, the depreciation of the French franc from 1922 to 1926. For the rest, he simply lists episodes during which exchange rates were flexible and asserts that in each case speculation was destabilizing. Even for the French episode, the evidence given by Nurkse does not justify any firm conclusion. Indeed, so far as it goes, it seems to me clearly less favorable to the conclusion Nurkse draws, that speculation was destabilizing, than to the opposite conclusion, that speculation was stabilizing." "The Case for Flexible Exchange Rates" in Essays in Positive Economics, p. 176, n. 9.

S. C. Tsiang offered a somewhat different conclusion: "Thus the instability of the French franc from 1923 to 1926 was really the result of an extremely elastic money supply which would have caused great instability of the economy whether the exchange rate was freely fluctuating or controlled." "Fluctuating Exchange Rates in Countries with Relatively Stable Economies," Staff Papers, October 1959, pp. 259-273.

2. Friedman, op. cit., p. 175.

3. Egon Sohmen, Flexible Exchange Rates (Chicago: University of Chicago Press, 1961), p. 50.

4. Ibid., p. 53.

The critics of flexible rate systems argue that the likelihood of destabilizing speculation increases the less frequently the market rate remains at a long-run equilibrium rate. They carry the argument further and maintain that any system of exchange rates is likely to be subject to sudden shocks. In a system of stabilized rates, the impact of the shock is met by a drain on reserves. In effect, the authorities speculate in the suspect currency by increasing their holdings of it in the hope that future events will be offsetting or that their policy measures will successfully restore payments equilibrium. The critics argue that speculation may become all one way if a similar shock occurs with a system of flexible rates.[1] Large speculative sales might then cause the currency to depreciate excessively and the increased cost of foreign exchange may complicate the problems of domestic economic policy.

The nature of speculation when changes in the equilibrium rate result from more rapid increases in prices in one country than in another is not generally discussed at the theoretical level.[2] A more rapid increase in prices in one country may result from the inability of its authorities to adopt successfully the measures to achieve price stability, or it may result from an explicit decision to forego greater price stability in order to seek a lower level of unemployment or a higher rate of economic growth. In both cases the authorities might have difficulty in maintaining a fixed rate for long; they might attempt to support the established rate through adoption and extension of exchange controls, and at times they might change the rate parity. Clearing the exchange market in this manner is an alternative to adopting a flexible exchange system.

Under a flexible exchange system, speculators might undertake exchange transactions in the expectation that the domestic price level may be higher in the future. Also, to the extent they discount the future, they may tend to cause the current exchange rate to deviate from the level justified by the prices and incomes of today. In these circumstances, the speculatively induced change in the exchange rate may have an independent upward effect on the domestic price level and thus tend to prove that the foresight of speculators was correct. But in the absence of speculation, the increase in the domestic price level might have been much smaller. Moreover, some of the expected events which are discounted may not come to pass; there may, for example, be a change in government policy. Nevertheless, any tendency of speculators to discount the future may have affected the exchange rate, and the higher cost of foreign exchange may have had its own effects on domestic price developments, and since prices and wages move upward with much greater ease than they decline, a higher price and wage

1. Sir Donald MacDougall, The World Dollar Problem (London: Macmillan & Co., 1957), p. 385.

2. One exception is S. C. Tsiang, "A Theory of Foreign-Exchange Speculation Under a Floating Exchange System," Journal of Political Economy, October 1958, pp. 399-418.

structure may become built-in even if the cost of foreign exchange subsequently falls. Where the price level is affected by speculatively induced movements in the exchange rate, it is not immediately clear how stabilizing and destabilizing speculation should be distinguished.

There is a measure of agreement that resolution of the controversy must be empirically based, since such an examination is necessary to point toward which sets of assumptions about speculative behavior seem to have the most general validity. But to the extent that history is a collection of special cases, it is likely to be exceedingly difficult, if not impossible, to obtain categorical answers. Within this limitation, this study examines the post-World War I experiences of a number of European countries with flexible rate systems for their relevance. This period provides a diversity of background for the operation of flexible rate systems because some countries had achieved relative price stability, while others were subject to varying amounts of monetary inflation.

This study attempts to answer a number of questions. Is there evidence that speculation was destabilizing of the overtracking variety? Did speculators reduce the amplitude of exchange rate fluctuations resulting from non-speculative transactions? Is there any evidence that speculators supported a currency in the face of economic and political shocks? Was the foresight of speculators sufficiently good so that it was unnecessary for the authorities to intervene in the exchange markets to keep rate movements orderly? To what extent did speculators exert an independent influence on the domestic economy through their impact on the level of the exchange rate? To what extent did speculators compromise the benefits of the additional degree of freedom for domestic policy claimed by advocates of a flexible rate system?

2. Some Problems of Definition, Measurement, and Interpretation

The disagreement in the evaluation of the nature of speculation is traceable to problems of definition, of measurement, and of interpretation. The distinction between stabilizing and destabilizing speculation noted earlier is inadequate because of its limited applicability to situations where the equilibrium rate changes through time. The measurement problem arises from the inadequacies of the data available for inferring how speculators have acted at different times. The problem of interpretation is that of attempting to separate the impact of speculative transactions from other economic forces which may have had similar impacts or offsetting impacts.

The Definitional Problem

For expository purposes, it is helpful to make two mutually exclusive assumptions about the exchange market in the context of a situation which suggests an unchanged equilibrium rate. Under one, changes in the domestic price level can be assumed to be independent of changes in the exchange rate; under the second, changes in the domestic price level can be assumed to be directly related to changes in the exchange rate. If the first assumption prevails, there is no reason for the actual rate ever to

deviate from the equilibrium rate. Any deviation arising from transactions of some individuals should result in immediate self-correcting transactions by other individuals who seek a speculative profit from buying the temporarily depressed currency in the expectation that it will soon appreciate to its equilibrium level. But if these expectations were shared by all, there would never be a reason for the actual rate to differ from the equilibrium rate. That any significant deviation from the equilibrium rate occurs suggests that some individuals do not accept the assumption of an unchanged equilibrium rate.

That the actual rate does differ from the equilibrium rate is sometimes attributed to the transactions of non-speculators. This distinction between speculators and non-speculators may be valid as long as changes in the domestic price level are assumed independent of changes in the exchange rate. When the assumption is dropped, as it must be if it is desired to move toward greater reality, the distinction loses its validity. In situations where changes in the exchange rate can affect the price level, individuals can only be considered non-speculators if their assets are perfectly hedged against changes in real value which might result indirectly from a change in the exchange rate. All other individuals qualify as speculators because the real value of their wealth will vary with changes in the exchange rate and in the price level. These induced price changes affect the real value of monetary assets, and they also affect the real value of the non-monetary assets of those individuals importantly involved in international trade. For example, those with large inventories of imported goods may suddenly find that the value of these goods in terms of domestic currency decline significantly if the price of foreign exchange declines. All individuals with a large volume of foreign transactions must be extremely cognizant of possible changes in the exchange rate; they have a strong financial incentive to time their transactions to take advantage of future rate movements.[1]

In the absence of a group of non-speculators, any deviation of the actual rate from an assumed equilibrium rate must result because some speculators do not believe that actual rates in the future will be the same as the current equilibrium rate. Transactions which increase the deviation of the actual rate from the equilibrium rate must be caused by speculators with one set of expectations about the future level of the rates, while the transactions which reduce these deviations must be undertaken by another group of speculators who have a different set of expectations. If an

---

1. The availability of forward market facilities for hedging the exchange risk attached to commitments to make payments abroad or to receive payments from abroad enables an individual to hedge a commitment once it is made. However, he may be able to make the same commitment at different times and he must seek to make it at the most propitious time to maximize his profits. The availability of forward facilities does not solve the problem of choosing the most favorable time to make the exchange commitment.

unchanged equilibrium rate is assumed, the transactions of the first set of speculators can be considered destabilizing, and those of the second set stabilizing.

The issue of whether speculators stabilize or destabilize the exchange market is thus a shadow issue; they do both. Deviations of the actual rate from the equilibrium rate occur because some individuals undertake transactions which are ex post deemed destabilizing, but these deviations do not become cumulative because other individuals undertake transactions which are ex post deemed stabilizing.[1]

The issues of substance then resolve into two questions as long as the assumption of an unchanged equilibrium rate is retained. The first is how extensive the deviations from the equilibrium rate are likely to become before the transactions of stabilizing speculators begin to dominate the exchange rate movement. The second is the extent to which these limitations on the rate deviation result from transactions by traders whose main business interests necessitates a large volume of exchange transactions, and the extent to which they result from the transactions of others who undertake exchange transactions solely for speculative profits. The transactions of this second group, who may be called investor-speculators, may dampen or amplify the rate fluctuations resulting from the transactions of the first group, the trader-speculators. If, as a group, the transactions of investor-speculators amplify the deviations from the equilibrium rate, then the progressive distortion in the relationship between the prices of domestic and foreign goods resulting from changes in the exchange rate will induce changes in imports and exports which tend to dampen further movements in the exchange rate.

When the domestic price level is not assumed to be independent of the price of foreign exchange, an increase in the price of foreign exchange resulting from either the transactions of speculators under a flexible rate system or the devaluation of a pegged rate will tend to raise domestic prices, both because imports will cost more and because the volume of exports and of domestic economic activity will be enhanced. An increase in domestic employment may lead to an increase in the price of foreign exchange because of a greater demand for imports, and thus lead to an even greater pressure on the domestic price levels. The upward pressure on domestic prices from any given change in the exchange rate will tend to be greater the smaller the volume of unemployed resources.

The equilibrium exchange rate is not unambiguous in the sense that only one rate is the equilibrium rate. Other things being equal, the equilibrium rate will differ with changes in income and employment, tastes, and resource availability. Thus there are many equilibrium exchange rates;

---

1. The dual character of speculation was recognized by N. Kaldor in "Speculation and Economic Stability," Review of Economic Studies, October 1939, pp. 1-27.

each set of government policies, geared to a particular level of employment or a particular level of domestic prices, has a particular equilibrium rate. In this connection, exchange speculators under a flexible rate system must obviously be cognizant of prospective changes in government policy. However, to the extent that their transactions influence the level of the actual rate, they influence the level of the equilibrium rate. This suggests a very stringent criterion for determining whether speculation is significantly destabilizing, which is whether it has an independent influence on the level of the equilibrium rate.

Moreover, the examination to determine whether speculation has an independent influence on the level of the equilibrium rate is an ex post one. The observer is able to view the period in retrospect while evaluating the quality of speculative foresight. The criterion for evaluating speculative foresight is not whether speculative transactions were those which a prudent man would have taken at a particular time, given the apparent constellation of political and economic factors, but whether speculative transactions had a significant influence on the level of the equilibrium rate and whether it caused deviations from this rate.

One aspect of the empirical examination is to determine the range of actual rate fluctuations in order to see whether there is a strong consensus among speculators as to future exchange rates. Another aspect is to determine whether investor-speculators undertake transactions which limit the deviations of the actual rate from the equilibrium rate, or whether their transactions increase these deviations. Still another is to determine whether speculative transactions have an independent influence on the level of the equilibrium rate.

The Measurement Problem

All of the techniques available for determining the nature of speculation and its impact on the exchange rate have their limitations. If reliance is placed on measuring short-term capital flows, problems of inadequate data arise. Estimates of these flows are frequently crude since they are derived as residuals, and the time periods involved generally are so long that offsetting movements within the period are washed out. Under a flexible rate system, there may be significant changes in the exchange rate even in the absence of any capital flows, simply as some participants in the market revise their expectations of future rates. Moreover, some speculative transactions may take the form of hoarding of foreign goods rather than foreign currencies so that no capital flow is apparent.

In many past instances, reliance has been placed on the Purchasing Power Parity technique to measure the impact of speculation. The PPP-technique was developed as a means of evaluating the impact of different rates of monetary inflation, with their resultant differential price level movements, on the exchange rate. The underlying assumption of this approach, when this technique is used to determine the impact of speculative forces, is that changes in the exchange rate which do not reflect

changes in the relative price levels and which cannot be explained by other changes in international supply and demand conditions are likely to be due to speculative forces. Thus deviations between the actual rate and the equilibrium rate inferred from comparable price indexes on the basis of the PPP-technique may result from speculative transactions.

One weakness of this approach is the lack of independence between speculatively induced changes in the exchange rate and the movements in the domestic price level. Where the foreign trade sector is large, the feedback from a speculatively induced change in the exchange rate to the domestic price level may be sufficiently strong and rapid to limit the appearance of a change in the deviation between the actual exchange rate and the derived equilibrium rate. Moreover, the PPP-technique is also handicapped by a limited number of observations, since only monthly price data is available.

What is needed is a direct indicator of changes in speculative anticipations over frequent, successive, short intervals so that changes in these anticipations can be compared with movements in the spot exchange rate, and with a wide array of other economic and political phenomena. Data on forward exchange rates, which are generally available either on a daily or weekly basis, can be organized to meet this need. Because the forward market provides the favorable opportunities for speculators who seek to profit from changes in the exchange rate, examination of the forward market data may lead to inferences on speculative anticipations at different times.

The differential between the forward and spot rates for a currency should generally reflect the money market interest differential between the domestic financial center and the foreign financial center.[1] If money market interest rates are higher abroad than at home, it is profitable for anyone with liquid balances to shift them abroad, and to cover the return shift against the exchange risk of a depreciation of domestic currency in terms of the foreign currency through a forward purchase of domestic currency. The shifts of arbitrage funds tend to cause the forward rate to move so as to reflect the money market interest differentials. If the supply of arbitrage funds were very elastic at a small positive yield, then the forward rate would change until the difference between the spot and forward rates almost fully reflected the interest differential, thus virtually eliminating any further incentive to shift funds.

To use the forward market data as a source of inferences about speculative behavior, it is important to show that in some cases the forward market

---

1. One of the earliest extended commentaries on the relation between the spot and forward rate is in John Maynard Keynes, A Tract on Monetary Reform, London, Macmillan & Co., 1924. See also Paul Einzig, The Theory of Forward Exchange, Macmillan & Co., 1937. For a comprehensive treatment of speculation in the forward market, see B. Reading, "The Forward Pound 1951-59," Economic Journal, June 1960, pp. 304-319.

is dominated by individuals who seek speculative profits from carrying exchange risk, rather than by commercial traders who use forward transactions to unload the exchange risk they have taken on in the course of their normal business transactions. Forward market facilities provide the opportunity to secure extremely high leverage since the margin requirements for forward transactions have been low. In some places the required margin has been fixed at 10 to 25 percent of the total amount of the transaction; elsewhere it has been nominal, set so as to cover any possible loss if the forward contract went into default because of an unexpected, adverse movement in the exchange rate. Traditionally, the leverage possibilities in the forward market for individuals who have sought to carry exchange risk have exceeded those available elsewhere.

The price which must be paid for leverage is the difference between the price of foreign exchange on the spot market and that on the forward market. In instances where the demand of individuals to take on exchange risk through forward transactions is limited, the price of leverage may be low or even negligible. As the demand for leverage increases, this price will tend to increase. As the spread between the forward rate and the spot rate begins to exceed the spread justified by the money market interest differentials, commercial traders are likely to find that using forward transactions to shuffle off the exchange risk associated with their foreign transactions becomes exceedingly expensive. In these circumstances, commercial traders can cover themselves against the exchange risk by combining a spot exchange transaction with a short term foreign loan or investment. For example, the importer with a commitment to make a payment in a foreign currency can buy this foreign exchange in the spot market, using his own funds or even borrowed funds, and invest the foreign exchange at short term abroad until the payment is due. Although this may be somewhat more inconvenient, the saving in costs makes it worthwhile. Thus, whenever the differential between the spot and forward rate becomes large, it is evident that forward transactions are dominated by those who are willing to pay the price.

The forward transactions of those who seek leverage alter the spread between the price of foreign exchange in the spot market and the forward market, as long as there is not an offsetting volume of arbitrage transactions. As long as the supply schedule of arbitrage funds is not perfectly elastic and as long as the schedule remains fairly stable, changes in the forward differential are likely to reflect changes in the demand for leverage.

By examining changes in the differential between the spot and forward rates, adjusted for changes in the money market interest differentials, it should be possible to see how speculative anticipations change through time. There are a number of empirical problems which must be considered: what money market interest rates should be used to adjust the forward-spot differential; how large the forward-spot differential, and changes in the differential, must be before they can be regarded as significant; and what part of the change in the differential should be attributed to changes in the

supply of arbitrage funds not due to changes in the money market interest differentials. While these issues can only be resolved in an arbitrary manner, the major conclusions about changes in speculative anticipations are not likely to be affected by the choice of assumptions provided two conditions prevail. The first is that changes in discount rates are representative of changes in short term rates; the second, that the non-monetary opportunity costs of funds used in arbitrage do not change abruptly or continuously. Both assumptions are plausible.

There remains the problem of how to interpret the adjusted forward differential. Several approaches are possible. At any moment, the level of the forward rate can be examined to determine whether speculators are bearish or bullish on a particular currency at that time. In particular, it may be helpful to examine their anticipations in the vicinity of turning points in the spot rate in order to infer whether they have amplified or dampened the cyclical fluctuations in the spot rate. Speculative anticipations, and changes in the anticipations, can also be related to other phenomena, for example, to whether a currency is undervalued, and becoming more or less so, and to political events, such as changes in governments and in governmental policies. Another approach is to examine the relation between changes in the forward rate and changes in the spot rate, which should lead to inferences on whether speculative transactions are consistent with the overtracking hypothesis.[1] These inferences must then be tested for their statistical significance.

In some cases the character of speculative behavior can be examined even without reference to the level of the equilibrium rate. For example, the overtracking hypothesis can be tested by examining whether forward speculators are bullish at the upper turning points, and bearish at the lower turning points. Since the level of the equilibrium rate is likely to change much more slowly than the rate in the spot market, it is not necessary to know the equilibrium rate to determine whether forward speculators cause the spot rate to overtrack.

The Purchasing Power Parity technique can be used to infer changes in the equilibrium rate. A major limitation of this technique is that it does not directly show changes in the equilibrium rate due to non-price changes,

---

1. Verification of this hypothesis does not settle the issue of whether speculators take their forward positions on the basis of the movement of the spot rate, or whether their forward transactions, undertaken in response to factors other than the movement in the spot rate, cause a reflected movement in the spot rate. Presumably speculators who follow both approaches may be active. Thus this distinction tends to become academic, since the events which lead some individuals to sell domestic currency forward are very likely to result in a depreciation of the currency in the spot market even in the absence of their forward transactions.

such as changes in income and in demand. Since changes in income and in these structural factors are not likely to be very rapid in relatively short periods, this approach is likely to provide a good proxy for changes in the equilibrium rate.

The forward market data can be usefully supplemented by an examination of the range of variation in the spot rate. The greater the consensus among participants in the exchange market about future exchange rates, and the longer this consensus holds, the smaller the range of fluctuations will be over any period. The range of fluctuations is a measure of the strength of the consensus. Within the context of an unchanged average actual rate, the range of daily variations may be large or small, depending on the strength of the consensus. The changes in the strength of the consensus as the average rate changes should indicate the extent to which speculators facilitate the process of rate adjustment and thus the extent to which the range of daily fluctuations provides a rough measure of the short-term supply of funds for stabilizing transactions.

The examination of the forward data should point to the behavior of those speculators most willing to take the large risks attached to a highly levered portfolio. Using the forward market data facilitates a distinction between the "pure" exchange speculators and other individuals, whose shifts of funds abroad to protect their real value from adverse changes in the price levels merit the appellation of flight capital. Some speculators may choose not to use the forward market, however, and thus it is desirable to examine the forward data together with other possible evidence of speculative behavior, including the information obtainable from the use of the Purchasing Power Parity technique and from the examination of the range of daily fluctuations.[1]

The Problem of Interpretation

The impact of speculatively induced changes in the exchange rate on the domestic economy complicates the interpretation of speculative behavior. If speculators wish to shift a sizeable volume of funds abroad, the domestic economy must generate an export surplus. Exports must increase relative to imports. The immediate impact of the increased demand for foreign exchange is to raise the price of foreign exchange in terms of domestic currency. The increased demand for exports, a result of the lower price of domestic goods in terms of foreign currency, is likely to exert a demand-pull impact on the domestic economy.

---

1. One limitation common to all analyses based on exchange market data is that the transactions of the bulls and bears are offset against each other, and so only a net speculative position is shown. To some extent, however, differences of opinion may still appear if the evidence suggests that speculators in the forward market are bullish, while those in the spot market are bearish.

Although the increased cost of foreign exchange may reduce the volume of imports, the higher cost of the remaining imports is likely to have a cost-push impact on the domestic economy. The adverse change in the terms of trade will tend to reduce real income, which also will have an inflationary impact on the domestic economy.

The resultant increases in the domestic price level from these several sources might prove that speculators were in fact correct in their anticipations, even though it is possible that the domestic price level would have remained relatively stable in the absence of the capital outflow. The situation is much more involved when there are domestic pressures which would tend to result in higher prices, even in the absence of speculative forces, and when these pressures become reinforced by the price-raising impacts of the capital outflow and the change in the exchange rate. In such situations it is necessary to attempt estimates of the strength of the domestic inflationary forces.

3. Flexible Exchange Rate Systems in Europe

The adoption of flexible exchange rate systems by most European countries after World War I followed immediately upon the cessation of American credits. During the war, most of the major currencies had been supported at levels slightly below their prewar dollar parities with the proceeds from U.S. private and government credits, the sale of gold and dollar securities, and exchange controls. Over the course of the war, however, prices within most European countries had increased from two hundred to four hundred percent or more of their 1913 levels, while U.S. prices had roughly doubled (see Table 1). Because of these differential price movements, some countries could have maintained their wartime exchange rates only with an extensive array of exchange controls once foreign credits were no longer available. When official U.S. credits to Great Britain were terminated in March 1919, the official support of sterling was withdrawn. Most other wartime belligerents, and many neutral nations, also ceased pegging their currencies at the wartime support levels.

Flexible rate systems were viewed as an interim measure until the various currencies could once again be stabilized at their prewar gold and dollar parities.[1] Because of the disparity of the movement of prices in some European countries and in the United States, the need for deflationary measures to secure this objective was generally recognized.

Six European countries, i.e., Great Britain, the Netherlands, Switzerland, Denmark, Sweden, and Norway, returned to the gold standard at their prewar parities between 1925 and 1927 (see Graph 1). Some other countries, i.e., Austria, Germany, Hungary, Poland, and Russia, eventually

1. Many non-European currencies also were not officially pegged; the Canadian dollar, the Japanese yen, and most Latin American currencies were in this group.

TABLE 1

Price Indexes in the United States and Selected European Countries, 1919–1926
(Annual Averages, 100 = 1913 or 1914)

	1919	1920	1921	1922	1923	1924	1925	1926	1927
Wholesale Prices									
United States	199	221	140	139	144	141	148	143	137
Belgium	-	-	-	367	497	573	558	744	847
France	364	520	352	334	428	499	561	718	630
Great Britain	258	307	197	159	159	166	159	148	141
Netherlands	304	292	182	160	151	156	155	145	148
Switzerland	-	-	196	168	180	176	161	148	147
Retail Prices or Cost of Living									
United States[1]	180	209	180	167	171	171	175	176	176
Belgium[2]	-	456	398	373	428	501	518	618	785
France[2]	268	371	337	301	332	380	425	554	556
Great Britain[1]	216	252	220	181	174	175	175	172	167
Netherlands[1]	-	219	208	187	173	173	179	171	167
Switzerland[1]	-	-	-	162	163	168	170	160	160

1. Cost-of-Living Index

2. Retail Price Index

Sources: European Currency and Finance, Federal Reserve Bulletin, International Abstract of Economic Statistics.

moved to stabilized rate systems after hyper-inflations necessitated currency reforms and the adoption of new currencies. An intermediate group of countries returned to a stable rate system at gold parities which were low in relation to their prewar parities. The gold value of the French franc after stabilization was twenty percent of its prewar value. For the Belgian franc this ratio was about fifteen percent and for the Italian lira, about twenty-seven percent.

The extent of official intervention in the exchange markets for the

TABLE 2

Government Expenditures and Fiscal Balance in Selected European Countries
(Millions of National Currency Units)

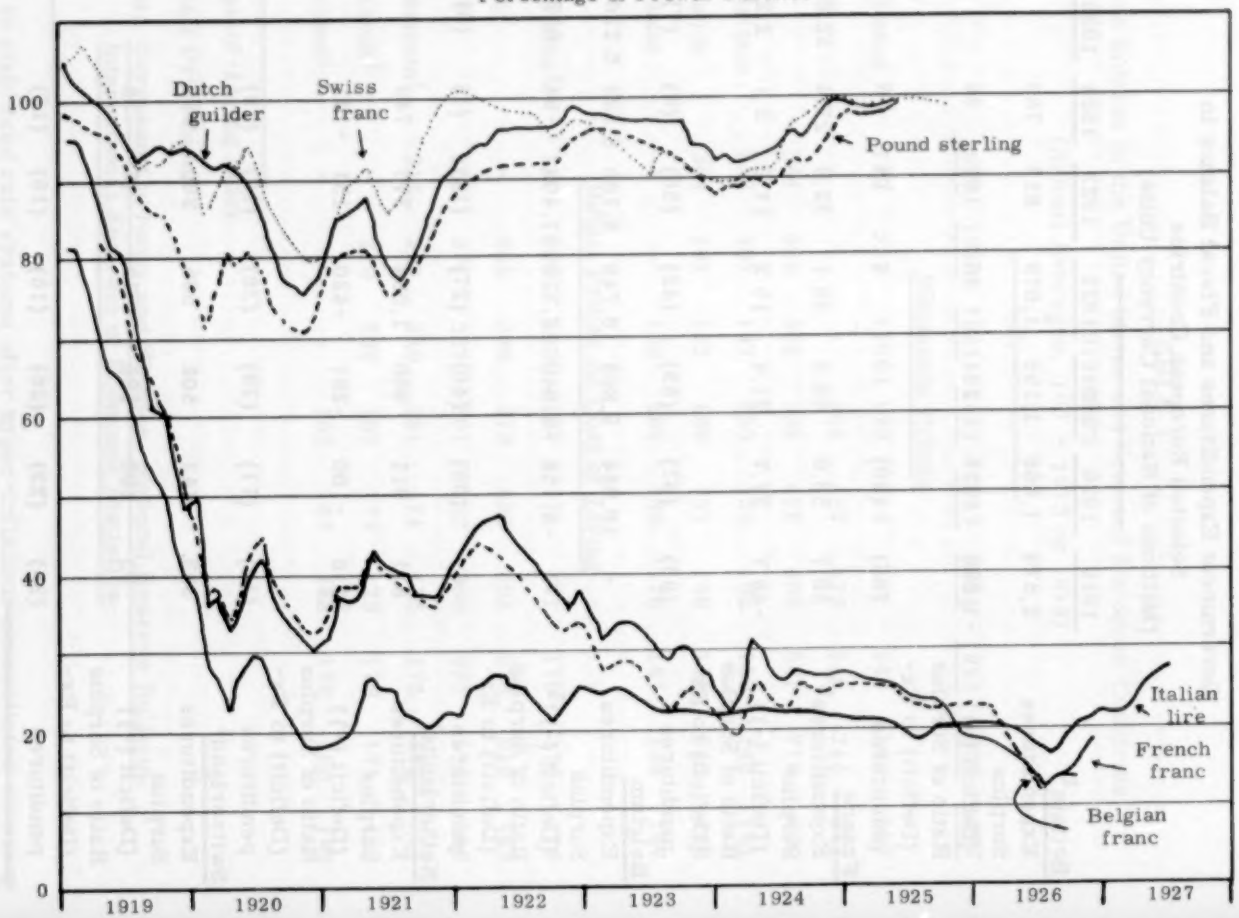
	1919	1920	1921	1922	1923	1924	1925	1926
Britain[1]								
Expenditures	2,579	1,666	1,195	1,079	813	789		
Surplus (Deficit (-))	-1,690	-326	231	46	102	48		
Ratio of Surplus (Deficit) to Expenditures	(66)	(20)	19	4	13	6		
France								
Expenditures[2]	58.7	56.0	40.2	36.1	32.8	27.4	32.0	41.3
Surplus (Deficit (-))	-48.7	37.7	21.4	15.3	11.9	2.1	2.2	3.2
Ratio of Surplus (Deficit) to Expenditures	(83)	(67)	(53)	(42)	(36)	(8)	(7)	8
Belgium								
Expenditures	-	10,944	9,808	8,749	8,186	8,322	8,220	
Surplus (Deficit (-))[3]	-	-4,158	-4,540	-2,323	-1,402	-84	-657	
Ratio of Surplus (Deficit) to Expenditures	-	(38)	(46)	(27)	(17)	(1)	(8)	
Netherlands[1]								
Expenditures	811	911	990	1,084	777	742		
Surplus (Deficit (-))	-116	-100	-281	-420	-151	-141		
Ratio of Surplus (Deficit) to Expenditures	(14)	(11)	(28)	(38)	(19)	(19)		
Switzerland[1]								
Expenditures	383	443	508	505	290	295		
Surplus (Deficit (-))	95	100	128	79	45	38		
Ratio of Surplus (Deficit) to Expenditures	(25)	(23)	(26)	(16)	(16)	(13)		

Note: The data is derived indirectly from official national sources. Nevertheless it should not be accepted as exact. It is sufficiently correct, however, to show whether there was a fiscal deficit or surplus, and whether it was relatively large or small.

1. European Currency and Finance.
2. The Public Finances of Postwar France.
3. The Monetary Experience of Belgium, 1914-1936.

Graph 1

Selected European Currencies as a Percentage of Prewar Parities



purpose of influencing the level or the rate of change of some currencies in terms of the dollar, or some other foreign currency, differed greatly. Both the Dutch and Swiss officials appear to have intervened extensively; the British not at all. In other countries the authorities intervened on occasion, at times to counter speculative attacks on their currency.[1]

Most of the West European currencies moved in the same direction at the same time in terms of the dollar during the early postwar years, depreciating together or appreciating together. They declined through 1919 and into the spring of 1920; they appreciated against the dollar for several months in the late spring of 1920, and then depreciated through most of the second half of 1920. After the end of 1921 there was less uniformity in their movements, although most of them, with the exception of the lira, became much weaker against the dollar in the last part of 1923. The similarity of the direction of movement, and their turning points, must be distinguished from the dispersion of their levels relative to their prewar parities.

The similarity in the turning points of the external value of the European currencies through 1922 is partly explainable in terms of events in the United States and world business conditions. The United States postwar industrial boom began in January 1919 and continued through March 1920; the industrial production index went from 85 to 120. The ensuing slump lasted until the beginning of 1922; the index of industrial production in 1921 was about 80. In general, other countries also participated in this postwar restocking boom at about the same time. Price movements in the United States had a similar time pattern; U.S. wholesale prices rose twenty percent from March 1919 to April 1920. In April 1920, the wholesale index was 247; a year later it was 145. Changes in U.S. prices were one of the important reasons for the swings in European exchange rates; the European demand for U.S. goods was then highly price inelastic because of the lack of alternative sources of supply, so that European currencies depreciated as dollar prices rose. When restrictive monetary measures were adopted in the United States in the spring of 1920, the supply of trade credit to European importers was greatly curtailed. The need to finance current imports on a cash basis, and to repay some of the outstanding credits, tended to depress the European currencies relative to the dollar.

Until 1921 or 1922 there was a large flow of U.S. capital into Europe.[2]

1. The apparent unwillingness of the authorities to intervene to make rate movements more orderly cannot be explained by a shortage of reserves, as the gold and foreign exchange holdings of most of the central banks were as large after the war as before. It may in part reflect a belief that intervention in the exchange market was not an appropriate function for a privately owned central bank. In some countries, however, central banks were allowed to deal in foreign exchange only at their prewar parities.

2. Hal Lary, The United States in the World Economy (Washington: U.S. Department of Commerce, 1943), pp. 93–94.

Some of it was trade credit extended to finance the sale of dollar goods, documented in dollars and therefore secure against exchange risk to the creditor. Another part of the capital inflow consisted of the purchase of monetary assets and physical assets in the countries with depreciated currencies. U.S. investors bought British assets and British investors bought assets in continental countries. These purchases were primarily undertaken in the expectation that the eventual return of depressed currencies toward their prewar parities would yield handsome profits. The belief that countries would be successful in re-attaining their gold parities was extremely strong, and dominated speculative expectations for several years.[1] Not until 1922, when the Austrian crown collapsed, was this faith badly shaken. Afterwards, European currencies became progressively subject to adverse speculative pressures.

The experiences of five countries with flexible rate systems, Britain, France, Belgium, the Netherlands, and Switzerland, are discussed in the next sections of this study. The success of these countries in achieving the appropriate restrictive policies differed (see Table 2). In Britain, the government maintained restrictive financial policies to bring sterling back to its gold parity; in France, in contrast, the absence of such policies negated the proclaimed efforts to attain its prewar parity. The Belgian experience is interesting because of the close links between the Belgian and French economies and the impact of these links on the ability of the Belgians to pursue policies indpendently of economic events in France. Although the financial policies of the Netherlands and Switzerland more nearly resembled those of France and Belgium than those of Britain, the Dutch guilder and the Swiss franc were eventually stabilized at their prewar parities.

IV. The Pound Sterling, 1919–1925

During most of World War I, the dollar rate for sterling was maintained by official support at $4.76, about $.10 below the prewar rate based on the gold parities. On March 19, 1919, official support was withdrawn and sterling became a flexible currency, subject only to a few remaining exchange controls which were removed before the end of the year. Sterling remained a flexible currency until April 30, 1925, when it was again stabilized at $4.86. During this sixty-one month interval, there is no evidence of any official intervention for the purpose of making rate movements more orderly, although the authorities entered the exchange market at times to accumulate dollars to make payments in the United States.

The return to prewar parity was one of the dominant objectives of postwar British economic policy. The first Interim Report of the Cuncliffe Committee of August 1919 recommended that the currency note issue be

1. The forward mark remained at a premium, indicating speculative bullishness about its future, until April 1922, when the spot mark had depreciated to about 6000 to the dollar from a prewar parity of 4 to the dollar.

reduced and that fiscal policy be designed to facilitate partial repayment of debt incurred by the government during the war. By fiscal 1920–21, the U.K. authorities had achieved a fiscal surplus.[1] After the postwar slump, which lasted through 1921, monetary measures were designed to achieve stable prices. It was hoped that relatively stable U.K. prices and an increase in U.S. prices would lead to an increase in the dollar price of sterling. The U.K. gold policy was designed to put upward pressure on U.S. prices by facilitating credit expansion in the United States; London maintained a free gold market for export sales, with the floor to the London price set by the U.S. buying price and sterling-dollar exchange rate. As sterling depreciated, buyers from countries whose currencies moved with sterling were priced out of the gold market, and gold flowed toward the United States.[2]

At the beginning of 1922, sterling was about $4.20, fifteen percent below its prewar parity; it was also relatively close to the Purchasing Power Parity rates suggested by both the wholesale and the cost-of-living indexes. At the same time, unemployment exceeded fifteen percent of the labor force, in contrast to prewar levels which ranged from two to four percent.[3] As much as half of the postwar unemployment was estimated to be in the export trades. Japan and India had taken over some export markets for cheap textiles formerly held by Britain, and Germany was furnishing coal as reparations to some of Britain's prewar customers. These adverse structural changes compounded the weakness in sterling which resulted from the deterioration in Britain's international investment position—its reduced holdings of U.S. and other foreign assets, and its greatly increased foreign debts.[4] Thus, despite the closeness between the actual rate and derived PPP rates, sterling was considerably overvalued at $4.20, in the sense that a sizeable reduction in money prices and money wages would have been necessary to achieve a level of unemployment as low as those of the prewar years at an equilibrium rate of $4.20.

The dilemma for monetary and fiscal policies was whether to pursue more restrictive policies in an attempt to quicken the return to parity at the cost of immediate greater unemployment or to follow expansionary policies as a solution for the high level of unemployment.

---

1. E. V. Morgan, Studies in British Financial Policy, 1914–1925 (London: Macmillan & Co., 1952), p. 375.

2. William Adams Brown, Jr., England and the New Gold Standard 1919–1926 (New Haven: Yale University Press, 1929), pp. 89–100.

3. International Abstract of Economic Statistics, 1919–1930 (London: International Conference of Economic Services, 1934), pp. 100–113.

4. A. C. Pigou, Aspects of British Economic History, 1918–1925 (London: Macmillan & Co., 1947), pp. 42–55.

If the former course had been adopted, the higher cost of sterling in terms of other currencies would have increased the unemployment in the U.K. export industries. If, instead, an expansionary policy had been adopted to alleviate the unemployment, the hope of a return to parity would have been scuttled. Larger imports would have been associated with higher domestic incomes, and this would have led to a decline in the sterling exchange rate. A decline in the exchange rate would undoubtedly have resulted in a cost-push element in domestic prices because of the higher sterling cost of imports. In view of this dilemma, the authorities adopted a passive waiting policy, and Britain remained in the economic doldrums.

Changes in the sterling-dollar rate after the end of 1921 were due to the very moderate changes in the relative price levels, to cyclical changes in income, and to capital movements, some of which were speculative. These moderate swings in the exchange rates were in contrast to the wide fluctuations of sterling in terms of the dollar in 1919–21.

The sterling-dollar rate was also subject to a set of special influences as a result of London's role as the major international center for foreign exchange transactions. Continental authorities bought or sold sterling to support their own currencies. On occasion, relatively large amounts of foreign-owned sterling balances were shifted to New York. Many shifts of funds from other European countries to the United States were routed through London; francs and guilders were initially converted into sterling and the sterling was then converted into dollars. When the continental demand for dollars in London was strong, sterling would tend to appreciate in terms of the continental currency while depreciating against the dollar. Thus, variations in the sterling-dollar rate resulted from London's role as an inter-market financial center as well as from economic and political events within the United Kingdom.

These financial disturbances to the sterling-dollar rate were compounded by competitive disturbances. At times, the currencies of some continental countries became extremely undervalued. As a result, U.K. exports to these countries tended to be priced out of the market, and U.K. exports to the United States tended to be undercut by cheaper continental sources.

The Forward Market Data

The forward market in London, which had been in operation before the war, developed rapidly in the face of exchange rate variations after withdrawal of the official support. The dollar price of sterling, both spot and forward, money market interest rates in New York and London, and the forward rate adjusted for net interest differentials and expressed as an annual percentage yield or cost, are shown in Graph 2.[1]

1. The data used in constructing these graphs have been obtained from Einzig's The Theory of Forward Exchange, and from The Manchester Guardian. The Guardian contains quotations on a day-to-day basis, for both spot and forward quotations, after April 1922.

Throughout the 1921-1925 period, with several significant exceptions, forward sterling remained at a significant premium in relation to the dollar, suggesting speculative anticipations of an appreciation of sterling.[1] The major exception occurred during the first eight months of 1922, when forward sterling was at a significant discount as sterling appreciated slightly from $4.20 to $4.40, after having climbed rapidly from $3.55 in the previous six months. As spot sterling continued to appreciate from $4.44 to $4.70 in the latter part of 1922 and into early 1923, speculative anticipations were reversed and forward sterling went to a significant discount.

The significant discount on forward sterling in several periods might be explained as a speculative response to several different factors. In the first part of 1922, sterling was somewhat overvalued by the wholesale indexes, and speculative forward sales of sterling might have been in reaction to the possible overvaluation. In the summer of 1922, sterling was undoubtedly affected by the collapse of the German mark and the recurrent crises over reparations. The discount on forward sterling in March 1924 coincides with the first major bear raid on the French franc. A large part of the flight from francs was into dollars, and the forward sterling rate reflected the shift of funds through London.[2]

The forward sterling rate was relatively resilient to the impact of political events within Britain. Perhaps the potentially most disruptive occurred in October 1923, when a government minister suggested the possible adoption of expansionary monetary and fiscal measures as a remedy for unemployment.[3] This statement induced a flight of funds from England; the forward sterling rate remained relatively unaffected. Parliament was dissolved in November 1923, and a Labour Government, which had campaigned in favor of a capital levy, was elected in December. This led to a renewal of the capital outflow, which continued until the government, in February 1924, affirmed that it would accept the principles of the Cuncliffe Report. Spot sterling declined from $4.55 to $4.30 during this three-month period of a large capital outflow; forward speculators, however, remained bullish on sterling.

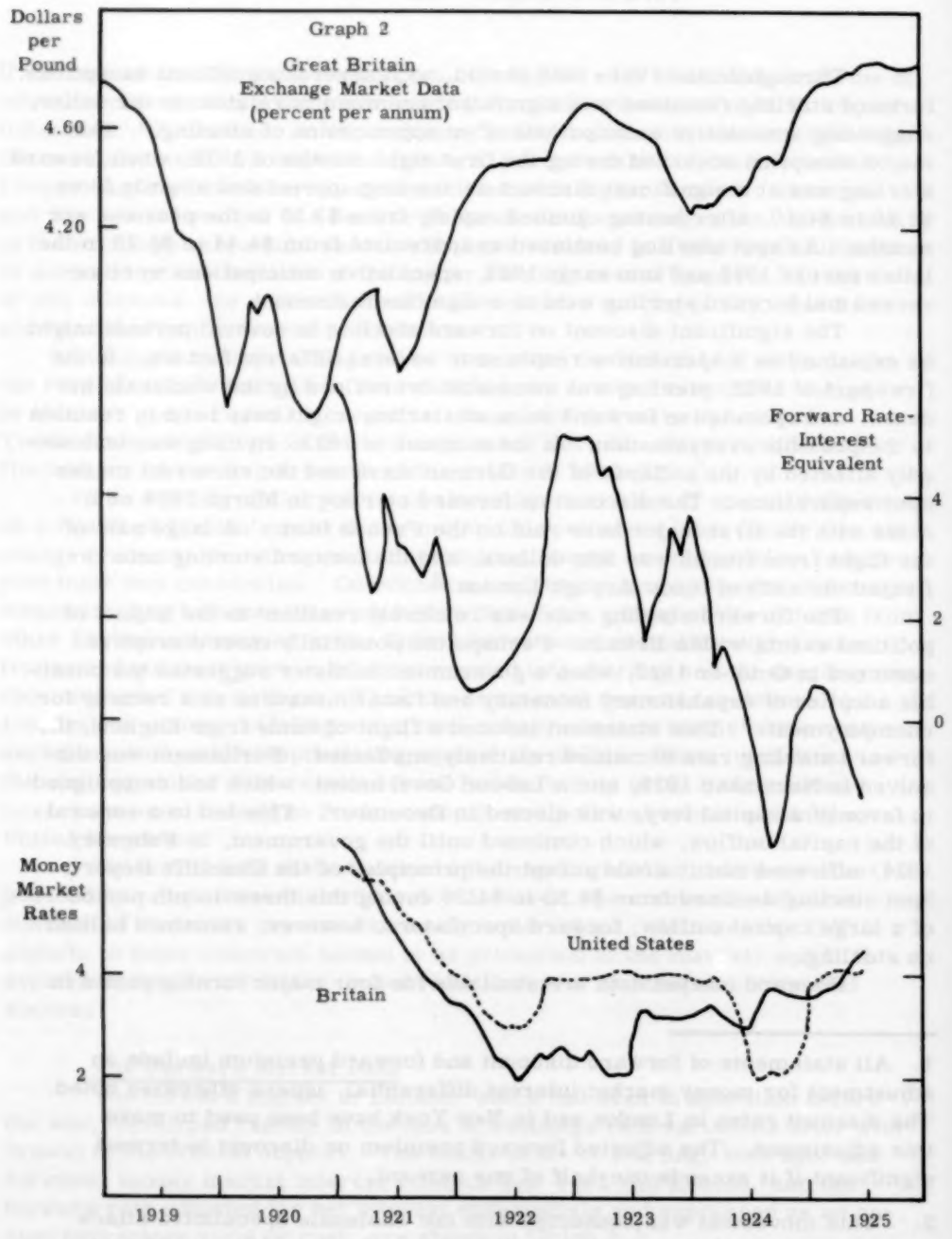
Forward market data are available for four major turning points in

---

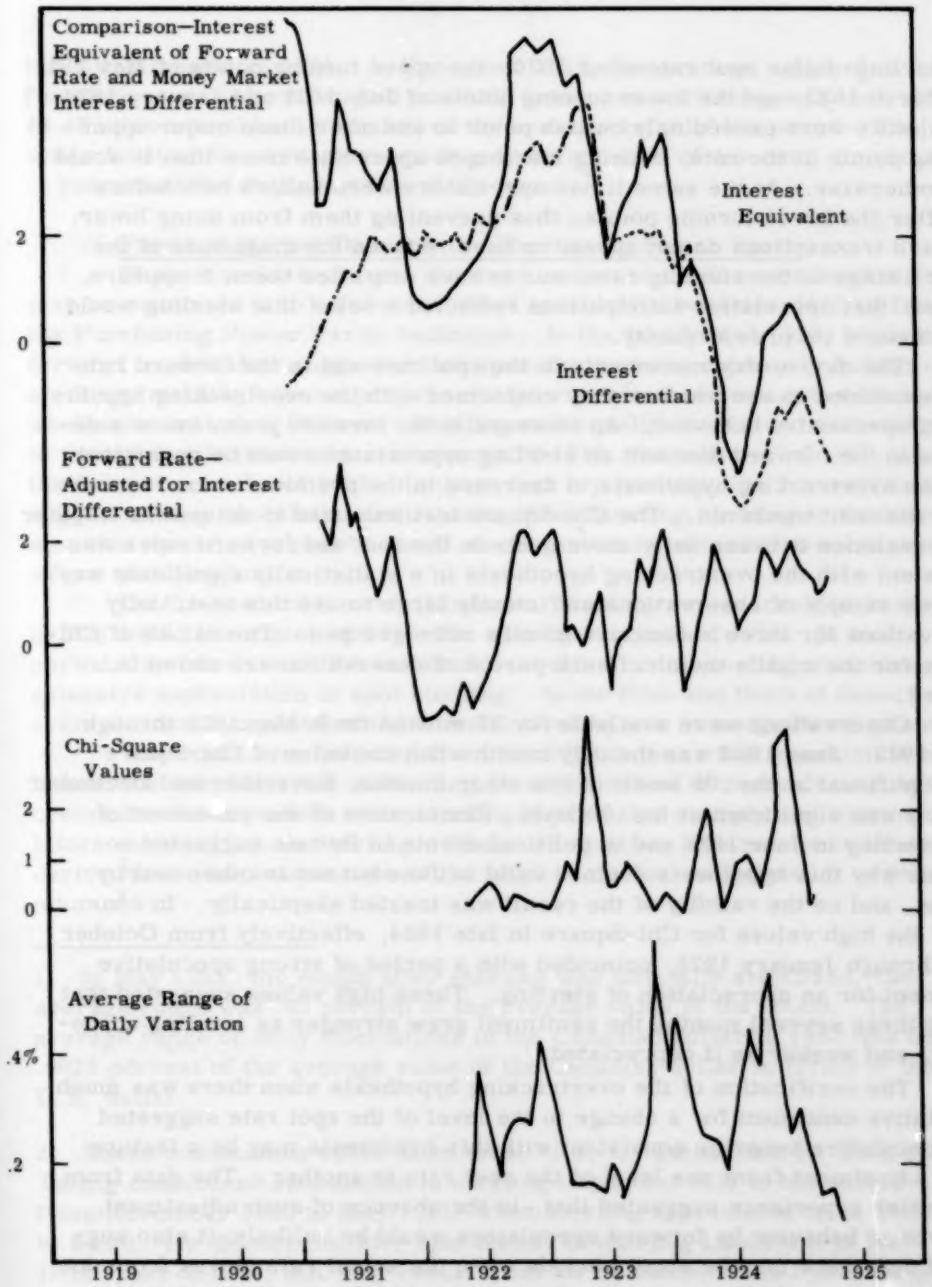
1. All statements of forward discount and forward premium include an adjustment for money market interest differential, unless otherwise noted. The discount rates in London and in New York have been used to make this adjustment. The adjusted forward premium or discount is termed significant if it exceeds one-half of one percent.

2. "This movement was connected with the wholesale speculative attack on the franc, which largely assumed the form of sales of forward francs against dollars. Much of these speculative transactions found its way to London . . . " Einzig, op. cit., p. 255.

3. Adams, op. cit., p. 25.


Graph 2
Great Britain
Exchange Market Data
(percent per annum)
Dollars per Pound
4.60
4.20
Forward Rate-Interest Equivalent
4
2
0
Money Market Rates
4
2
United States
Britain
1919
1920
1921
1922
1923
1924
1925

Graph 2 — cont.


Comparison—Interest Equivalent of Forward Rate and Money Market Interest Differential
Interest Equivalent
Interest Differential
2
0
Forward Rate—Adjusted for Interest Differential
2
0
Chi-Square Values
2
1
0
Average Range of Daily Variation
.4%
.2
1919
1920
1921
1922
1923
1924
1925

the sterling-dollar spot rate after 1920: the upper turning points of May 1921 and March 1923; and the lower turning points of July 1921 and January 1924. Speculators were exceedingly bullish prior to and after these major upper turning points in the rate, causing sterling to appreciate more than it would have otherwise. At the same time, speculators were bullish both before and after the lower turning points, thus preventing them from being lower. Forward transactions do not appear to have reduced the magnitude of the major swings in the sterling rate, nor to have amplified them; it appears, instead, that speculative anticipations reflected a belief that sterling would move toward its prewar parity.

The day-to-day movements in the spot rate and in the forward rate were examined to see whether they conformed with the overtracking hypothesis of speculative behavior. An increase in the forward premium or a decrease in the forward discount as sterling appreciated would be consistent with the overtracking hypothesis; a decrease in the premium or an increase in the discount would not. The Chi-Square test was used to determine whether the correlation between daily movements in the spot and forward rates was consistent with the overtracking hypothesis in a statistically significant way. To get a sample of observations sufficiently large to use this test, daily observations for three successive months were grouped. The values of Chi-Square for the middle month of each period of observation are shown in Graph 2.

Observations were available for 37 months from May 1922 through April 1925. June 1923 was the only month when the value of Chi-Square was significant at the .95 level; in two other months, November and December 1924, it was significant at the .90 level. Examination of the movement of spot sterling in June 1923 and in political events in Britain suggested no reasons why this hypothesis seemed valid in June but not in other nearby months, and so the validity of the result was treated skeptically. In contrast, the high values for Chi-Square in late 1924, effectively from October 1924 through January 1925, coincided with a period of strong speculative sentiment for an appreciation of sterling. These high values suggested that during these several months the sentiment grew stronger as sterling appreciated, and weaker as it depreciated.

The verification of the overtracking hypothesis when there was much speculative sentiment for a change in the level of the spot rate suggested that speculative behavior consistent with this hypothesis may be a feature of the adjustment from one level of the spot rate to another. The data from the British experience suggested that, in the absence of such adjustment, this type of behavior by forward speculators would be unlikely; it also suggested that there may be some periods when the actual rate moves from one level to another which may not be characterized by forward speculation which conforms with the overtracking hypothesis.

The average range of daily fluctuations in the sterling-dollar rate tended to increase greatly in the period when the average level of the spot rate was changing rapidly, as in November and December of 1923, and in

the latter months of 1924. In November 1923, the average range of daily fluctuations was about two and one-half cents; since sterling was then about $4.40, this was about .6 percent of the average value of the pound. The lower limit of the average range was about one-half of one cent, or about .1 percent of the average value of the pound. [1]

Forward Market Data and the Purchasing Power Parity

These conclusions from the examination of the forward market data may be compared with inferences about speculative behavior derived from the Purchasing Power Parity technique. In the last half of 1924, when the forward market data suggested that speculators were extremely bullish on sterling, sterling became progressively overvalued according to both the wholesale and retail indexes. Sterling was equally overvalued according to the wholesale price indexes in much of 1919, the summer of 1920, and the first six months of 1921. In these last two periods, the forward data suggest strong speculative bullishness about sterling, leading to the conclusion that speculative purchases of sterling must have tended to increase the extent of overvaluation.

On the basis of the retail indexes, sterling was overvalued in the last half of 1920, the winter of 1921–22, and the last half of 1924. In each period, the increase in the extent of overvaluation of sterling coincides with extensive appreciation of spot sterling. In the first and third of these periods, overvaluation increased while forward sterling was at a significant premium.

The conclusion inferred from the PPP rates is that speculation enhanced the overvaluation of sterling since the extent of overvaluation increased in those periods when sterling appreciated sharply and when speculators were bullish on sterling. These conclusions thus reinforce those derived from the examination of the level of the forward discount or premium. [2]

---

1. In May 1925, the month after sterling was officially stabilized, the average range was .05 percent of the average value of the pound. The average range of daily fluctuations in the Canadian dollar in 1960 was below .0025 percent of the average value of the Canadian dollar in terms of the U.S. dollar.

2. These conclusions differ from those of S. C. Tsiang and W. Stolper. Tsiang considered speculation in sterling important only in the period from November 1920 to May 1921, when sterling appreciated from $3.35 to $4.20. He did not consider speculation in sterling important in the latter part of 1924; this may result from his reliance on the Purchasing Power Parity technique using only wholesale price indexes. See S. C. Tsiang, "Fluctuating Exchange Rates in Countries with Relatively Stable Economies," p. 256.

Stolper concluded that deviations between the actual rate and the derived PPP rates could in most periods be satisfactorily explained by

Conclusion

Through most of the interval from March 1919 through April 1925, speculators in the forward market tended to support sterling. The support given sterling at its lower turning points reflected speculative anticipations of the eventual return of sterling to its prewar parity but, since forward speculators were equally bullish on sterling at its upper turning points, it cannot be concluded that their transactions dampened the major swings in sterling. The only evidence of forward speculation consistent with the overtracking hypothesis was in the last several months of 1924, when the expectations of a return to parity were very strong. In other months, when the average dollar price of sterling changed relatively little, the average range of daily fluctuations was quite small; traders and others were apparently sufficiently confident of the expected level of sterling in the near future to time their transactions in a manner which prevented large deviations from recent rates.

The upward pressure on sterling from speculation, particularly during the second half of 1924, complicated the problem of monetary management. Because of the overvaluation of sterling resulting from speculative factors, U.K. exports and U.K. employment were lowered by the relatively higher price of U.K. goods in foreign markets. In late 1924 and early 1925, speculation forced sterling toward its prewar parity before the underlying price and cost situation warranted stabilization at this level. The authorities then faced the dilemma of whether to ratify this decision at the possible cost of having to adopt restrictive monetary policies to support the rate, or to delay their de facto stabilization. As long as speculators believed that the proclaimed goal of the authorities might be achieved soon and increased their purchases of sterling, the authorities could only have reduced the rate by direct intervention in the exchange market, or through an easier money policy which would have reduced the interest incentive to switch arbitrage funds to Britain. Lower interest rates, however, might have hindered the attainment of stable domestic prices.

Speculators were able to force the hand of the authorities because the authorities had proposed the return to the prewar parity as their objective, but were unable or unwilling to take the measures to accomplish

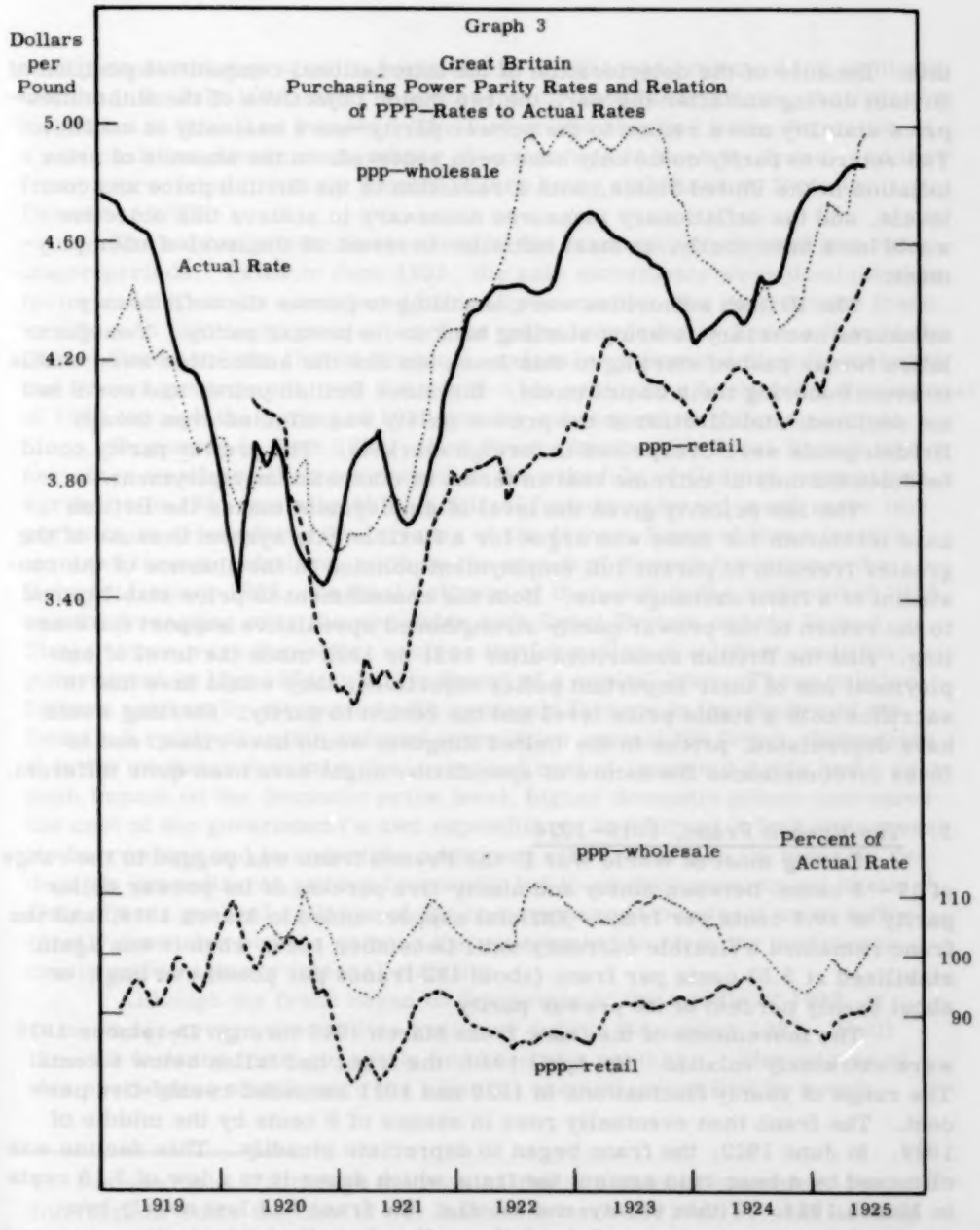
---

changes in business conditions in the United States and the United Kingdom, except for a period in early 1924. At that time, he maintained, the decline in activity in the United States and the increase in the activity in the United Kingdom should have led to a depreciation of sterling, whereas it actually appreciated. In part, the strengthening of sterling during this period reflected the return flow of funds to the United Kingdom after the Labour Government announced that its policies would be guided by the principles of the Cuncliffe Report. W. Stolper, "Purchasing Power Parity and the Pound Sterling from 1919–1925," Kyklos, Fasc. 2, 1928, pp. 240–269.

Graph 3

Great Britain

Purchasing Power Parity Rates and Relation of PPP–Rates to Actual Rates



this. Because of the deterioration of the international competitive position of Britain during and after the war, the two major objectives of the authorities—price stability and a return to the prewar parity—were basically in conflict. The return to parity could only have been achieved, in the absence of price inflation in the United States, with a reduction in the British price and cost levels, and the deflationary measures necessary to achieve this objective would have been costly, at least initially, in terms of the level of unemployment.

The British authorities were unwilling to pursue the deflationary measures necessary to bring sterling back to its prewar parity. Yet speculative forces pushed sterling to this level, so that the authorities were unable to avoid honoring their commitment. But since British prices and costs had not declined, stabilization at the prewar parity was effected even though British goods were overpriced in foreign markets. The prewar parity could be defended only at extreme cost in terms of domestic unemployment.

The low priority given the level of employment makes the British case irrelevant for those who argue for a flexible rate system because of the greater freedom to pursue full employment policies in the absence of the constraint of a fixed exchange rate. Both the commitment to price stability and to the return to the prewar parity strengthened speculative support for sterling. Had the British authorities after 1921 or 1922 made the level of employment one of their important policy objectives, they would have had to sacrifice both a stable price level and the return to parity. Sterling would have depreciated, prices in the United Kingdom would have risen, and in these circumstances the nature of speculation might have been quite different.

5. The French Franc, 1919–1926

During most of World War I, the French franc was pegged in the range of 17–18 cents, between ninety and ninety-five percent of its prewar dollar parity of 19.3 cents per franc. Official support ended in March 1919, and the franc remained a flexible currency until December 1926, when it was again stabilized at 3.93 cents per franc (about 122 francs per pound sterling), or about twenty percent of its prewar parity.

The movements of the franc from March 1919 through December 1926 were extremely volatile. By April 1920, the franc had fallen below 6 cents. The range of yearly fluctuations in 1920 and 1921 exceeded twenty-five percent. The franc then eventually rose in excess of 9 cents by the middle of 1922. In June 1922, the franc began to depreciate steadily. This decline was climaxed by a bear raid against the franc which drove it to a low of 3.15 cents in March 1924. Within twenty-one months, the franc had lost nearly two-thirds of its value in terms of the dollar. Nearly half of this decline occurred within twenty weeks, for the franc had sold for nearly 6 cents at the end of October 1923.

The franc remained relatively stable in terms of the dollar during the last eight months of 1924, at a level slightly in excess of 5 cents; however, it declined in terms of sterling as sterling appreciated. A second period of

nearly continual depreciation began early in 1925 and continued for eighteen months. By July 1926, after a succession of confidence crises, the franc had fallen to less than 2 cents. The last crisis resulted in the formation of a new coalition government, whose objective was to save the franc, and the franc appreciated to nearly 4 cents before being stabilized de facto in December 1926.

The exchange rate fluctuations of the franc can be divided into two major periods. Prior to June 1922, the rate movements were dominated by factors which can be considered basically economic: the need for the franc to adjust to relatively greater wartime price increases in France than in the United States and the United Kingdom, extensive postwar deficit financing by the French government, and the world-wide restocking boom from the summer of 1919 through the spring of 1920 and the subsequent world-wide slump.

Beginning in June 1922, when France broke with Great Britain and her other wartime allies over whether the schedule of German reparations agreed to in 1921 was alterable, political factors assumed much more importance in determining the fortunes of the franc. Some of these factors involved foreign relations, such as the march of French troops into the Ruhr in January 1923, the final collapse of the mark in the summer of 1923, and the wrangling over the war debts with Great Britain and the United States; some were domestic, such as the formation of a leftist coalition government in May 1924, with its threat of a capital levy. These political factors continually interacted with economic factors in the decline of the franc. A political crisis induced speculation against the franc, depressing it in the exchange markets; the increased cost of imported goods had a cost-push impact on the domestic price level; higher domestic prices increased the cost of the government's own expenditures and forced it both to increase its borrowings and to undertake additional efforts to raise its taxes or reduce its expenditures, which frequently led to another government crisis.[1] Only after a general collapse of the franc appeared imminent were sufficiently extensive powers granted the government to stem the confidence crisis and to achieve internal stability.

Although the franc began to depreciate in the middle of 1922, it came under severe speculative pressure only in November 1923. Until then the forward franc generally had been at a premium. The origins of this speculative attack were not clear; it seemed unjustified by events

1. For example, "Although the rate on Paris continues unsettled, the fluctuations since the date of our last issue have not been so wide as they were in the previous week. This is due no doubt in part to the fact that transactions have been on a smaller scale, speculators having preferred to sit on the fence and await developments at the Quai d'Orsay and at the Wilhelmstrasse." Samuel Montagu & Co., Weekly Review of Foreign Exchanges, May 29, 1924, p. 2. Such statements appeared almost weekly in the contemporary journals.

within France.[1] Some contemporary accounts maintained that the attack began in Paris as foreign owners of francs began to sell them for foreign currencies; others, that it began in Amsterdam or Vienna as the financial manipulators who had profited so well from their machinations in the mark and the schilling looked elsewhere for profits.[2]

After feeble attempts to use discount policy and to strengthen their controls against capital exports, the French authorities adopted a more stringent domestic stabilization program, including a twenty percent increase in all taxes, and countered the pressure on the franc by intervening in the exchange market. Massive intervention financed by a $125 million credit from foreign banking firms turned the tide, many speculators reversed their transactions, and within a month the franc climbed from 3.15 cents to over 6 cents. This experience demonstrated that the value of the franc could be greatly affected by the speculative transactions in the face of weak or impotent government policies, and the franc remained under extreme speculative suspicion until it was eventually pegged.

The Forward Market Data

The withdrawal of official support of the franc in March 1919 and the rapid movements in the exchange rate provided a great stimulus to the development of forward exchange operations in France.[3] The forward franc, adjusted for interest differentials, was at a significant premium against sterling through most of 1921 and 1922, and from March through October of 1923 (see Graph 4).[4] In 1922, it was at a significant discount only in November,

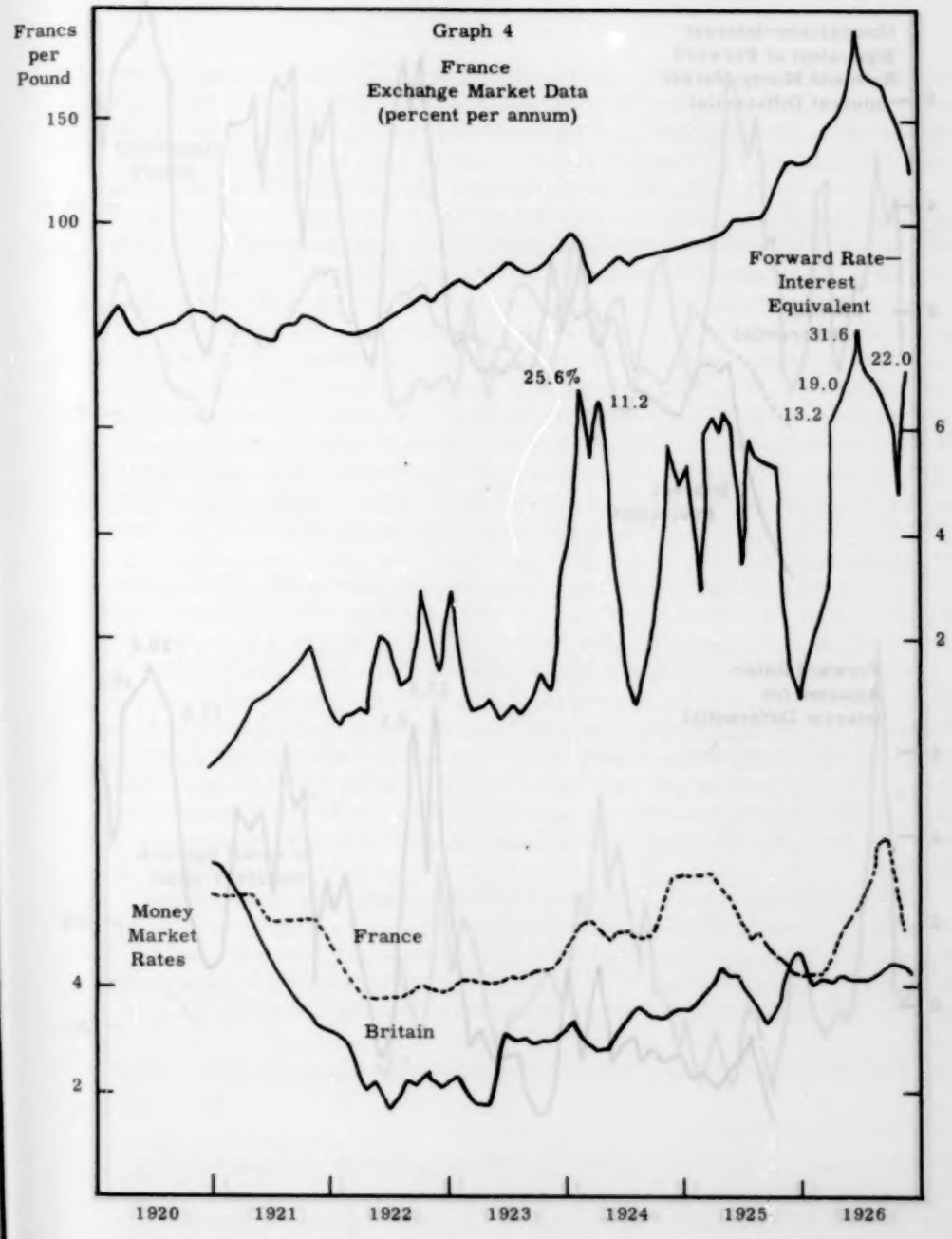
---

1. Some time in November 1923, it became known that the Treasury would not be able to fully honor its 1921 convention with the Bank of France that its borrowings from the Bank would be reduced by 2 billion francs annually. In 1922, this convention had also not been honored, without any apparent adverse effects.

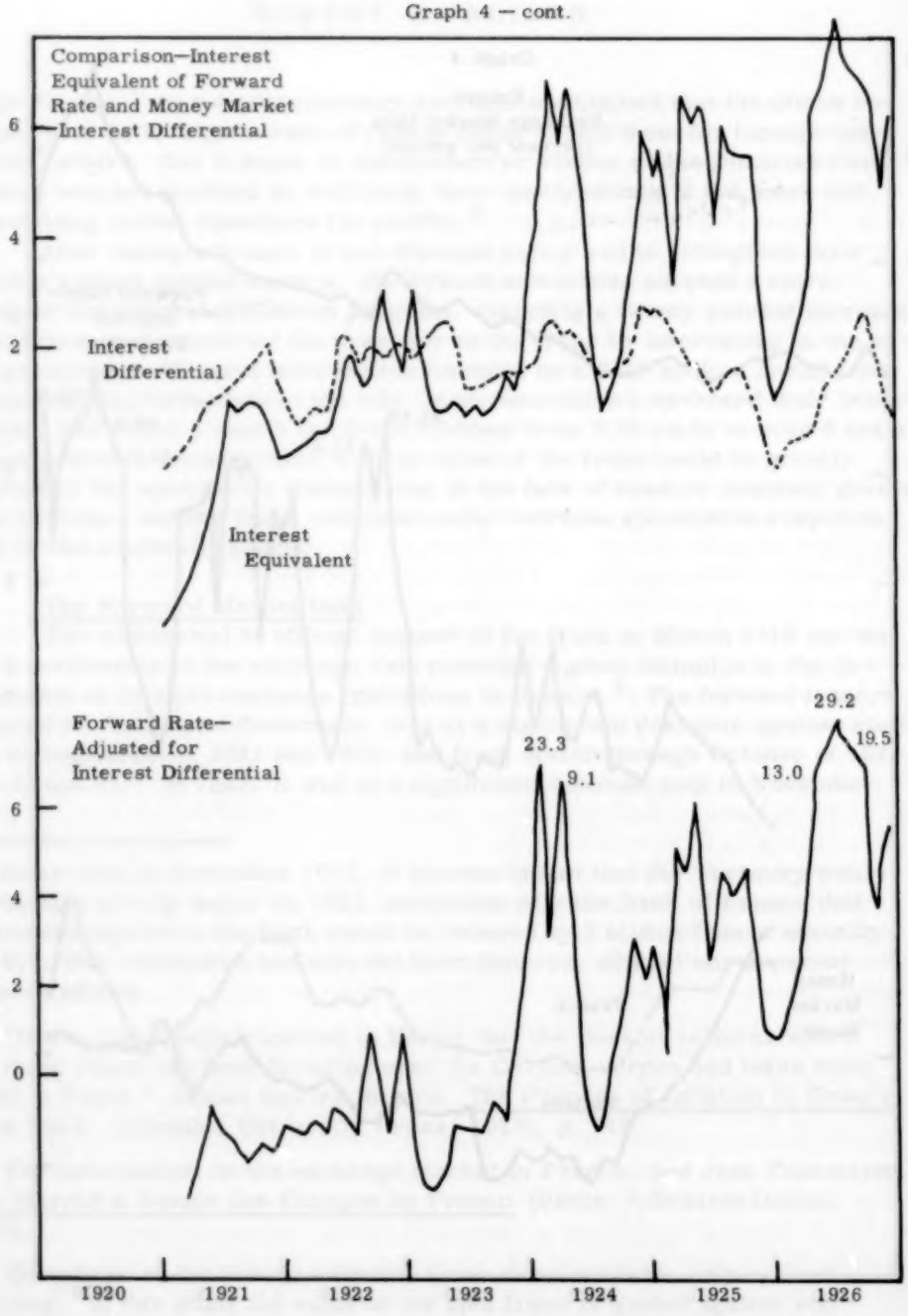
2. "It was continually rumored in Berlin that the flock of vultures which for three years had been hovering over the German corpse had taken their flight to Paris." James Harvey Rogers, The Process of Inflation in France (New York: Columbia University Press, 1929), p. 125.

3. For information on the exchange market in France, see Jean Casamajor, Les Marché à Terme des Changes en France (Paris: Librairie Dalloz, 1925).

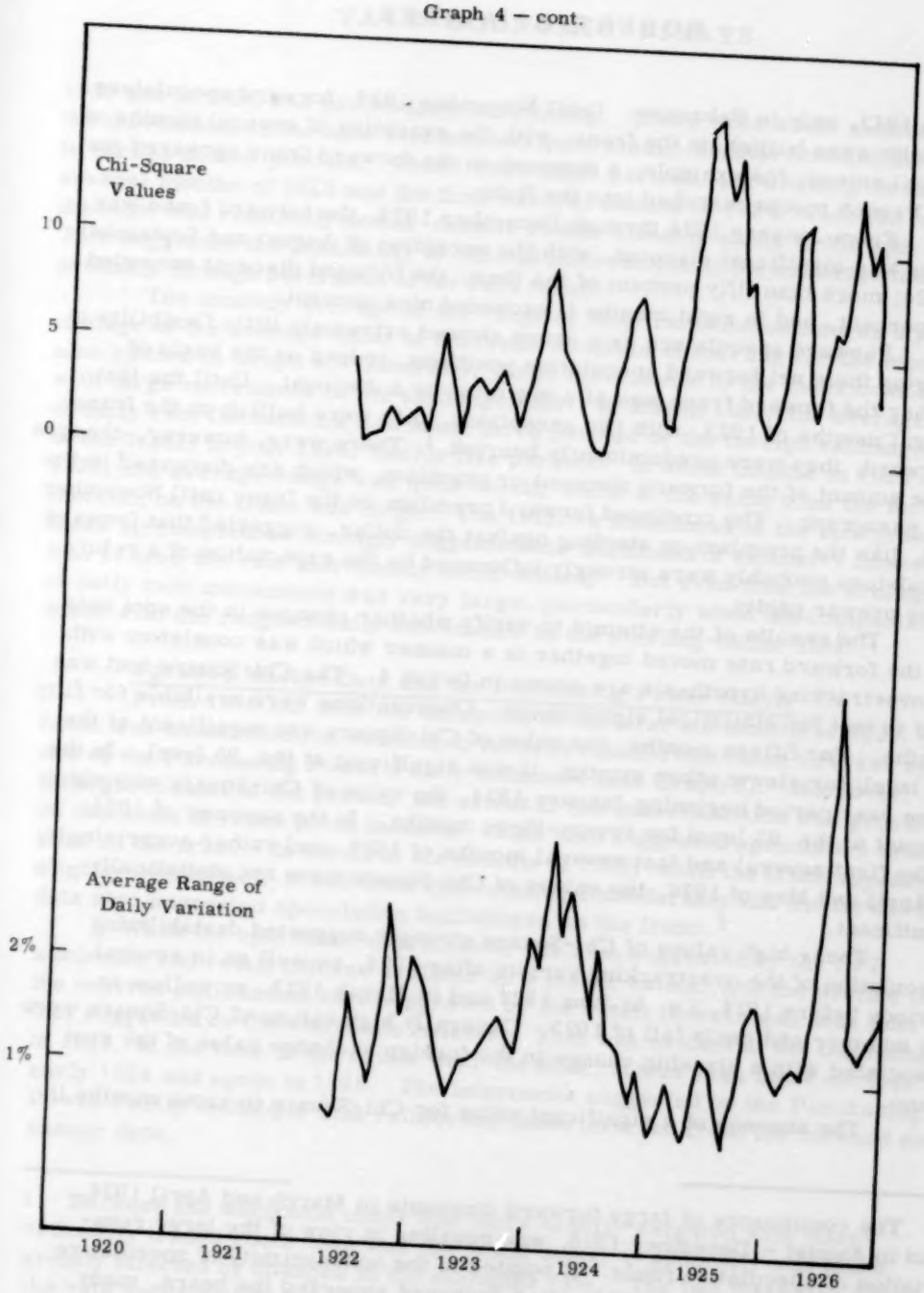
4. Quotations on the forward French franc were available only against sterling. In this study the value of the spot franc is quoted against sterling, although at times its value is given in terms of the dollar to show the extent of fluctuations in relation to its prewar parity better. Since the quotation against sterling is in terms of the number of francs necessary to buy a pound, an upper turning point means a minimum value for the franc in terms of sterling.


Graph 4
France
Exchange Market Data
(percent per annum)
Francs per Pound
150
100
Forward Rate—Interest Equivalent
31.6
22.0
19.0
13.2
25.6%
11.2
6
4
2
Money Market Rates
France
Britain
4
2
1920
1921
1922
1923
1924
1925
1926

Graph 4 — cont.


Comparison—Interest Equivalent of Forward Rate and Money Market Interest Differential
6
4
2
Interest Differential
Interest Equivalent
Forward Rate—Adjusted for Interest Differential
23.3
9.1
13.0
29.2
19.5
6
4
2
0
1920
1921
1922
1923
1924
1925
1926

Graph 4 — cont.


Chi-Square Values
10
5
0
Average Range of Daily Variation
2%
1%
1920
1921
1922
1923
1924
1925
1926

and in 1923, only in February. Until November 1923, forward speculators generally were bullish on the franc, with the exception of several months of political crises; for example, a discount on the forward franc appeared just after French troops marched into the Ruhr.

From January 1924 through December 1926, the forward franc was always at a significant discount, with the exception of August and September of 1924; more than fifty percent of the time, the forward discount exceeded four percent, and in eight months it exceeded nine percent.

Forward speculators as a group showed extremely little flexibility in changing their net forward speculative positions, judged on the basis of whether the forward franc was at a premium or a discount. Until the last several months of 1923, with two exceptions, they were bullish on the franc; afterward, they were predominantly bearish.[1] There were, however, changes in the amount of the forward discount or premium, which are discussed in the next paragraph. The continued forward premium on the franc until November 1923, like the premium on sterling against the dollar, suggested that forward speculators probably were strongly influenced by the expectation of a return to the prewar parity.

The results of the attempt to verify whether changes in the spot rate and the forward rate moved together in a manner which was consistent with the overtracking hypothesis are shown in Graph 4. The Chi-Square test was used to test for statistical significance. Observations were available for fifty months. For fifteen months, the value of Chi-Square was significant at the .99 level; for eleven other months, it was significant at the .95 level. In the three year period beginning January 1924, the value of Chi-Square was significant at the .95 level for twenty-three months. In the summer of 1924, in the first several and last several months of 1925, and rather surprisingly, in April and May of 1926, the values of Chi-Square were not statistically significant.

These high values of Chi-Square strongly suggested destabilizing speculation of the overtracking variety after 1924, as well as in several periods before 1924, i.e. in June 1922 and in March 1923, as well as in the summer and early fall of 1923. Generally high values of Chi-Square were associated with a sizeable change in the foreign exchange value of the spot franc.

The absence of a significant value for Chi-Square in some months in

---

1. The continuance of large forward discounts in March and April 1924, and in August – December 1926, was puzzling in view of the large repatriation of speculative funds. It pointed to the nonunanimity of speculative anticipations. In March 1924, after Poincaré squeezed the bears, many speculators were desperately short of francs, and apparently were forced to pay a large discount to renew their transactions. Also, many were anticipating that the government's intervention would fail because of insufficient funds. See Einzig, op. cit., p. 281.

1925 and in 1926 was particularly surprising. There were some months when the movements in the spot rate seemed quite small, despite forward discounts in excess of four percent. There were other periods, particularly in the last several months of 1925 and the first several months of 1926, when the forward discount was relatively small, despite rather sizeable shifts in the spot rate. This suggested the possibility of official intervention in the exchange markets, probably through purchases of forward francs.

The monthly average of the range of daily rate fluctuations as a percentage of the average value of the franc in terms of sterling for the month is also plotted in Graph 4. Generally, large movements in the index coincided with large movements in the exchange rate. In March 1924, the average range of daily rate fluctuations was about three percent of the foreign exchange value of the franc; in July 1926, nearly five percent. In some months in 1925 and 1926, the average range was quite small, while at the same time the forward discount on the franc was large. The relative steadiness of the spot franc in these circumstances strongly suggested the likelihood of extensive intervention to keep the rate movements more orderly. But even then the average range of daily rate movements was very large, particularly when the comparison is made with the range of daily movements in the sterling-dollar rate.

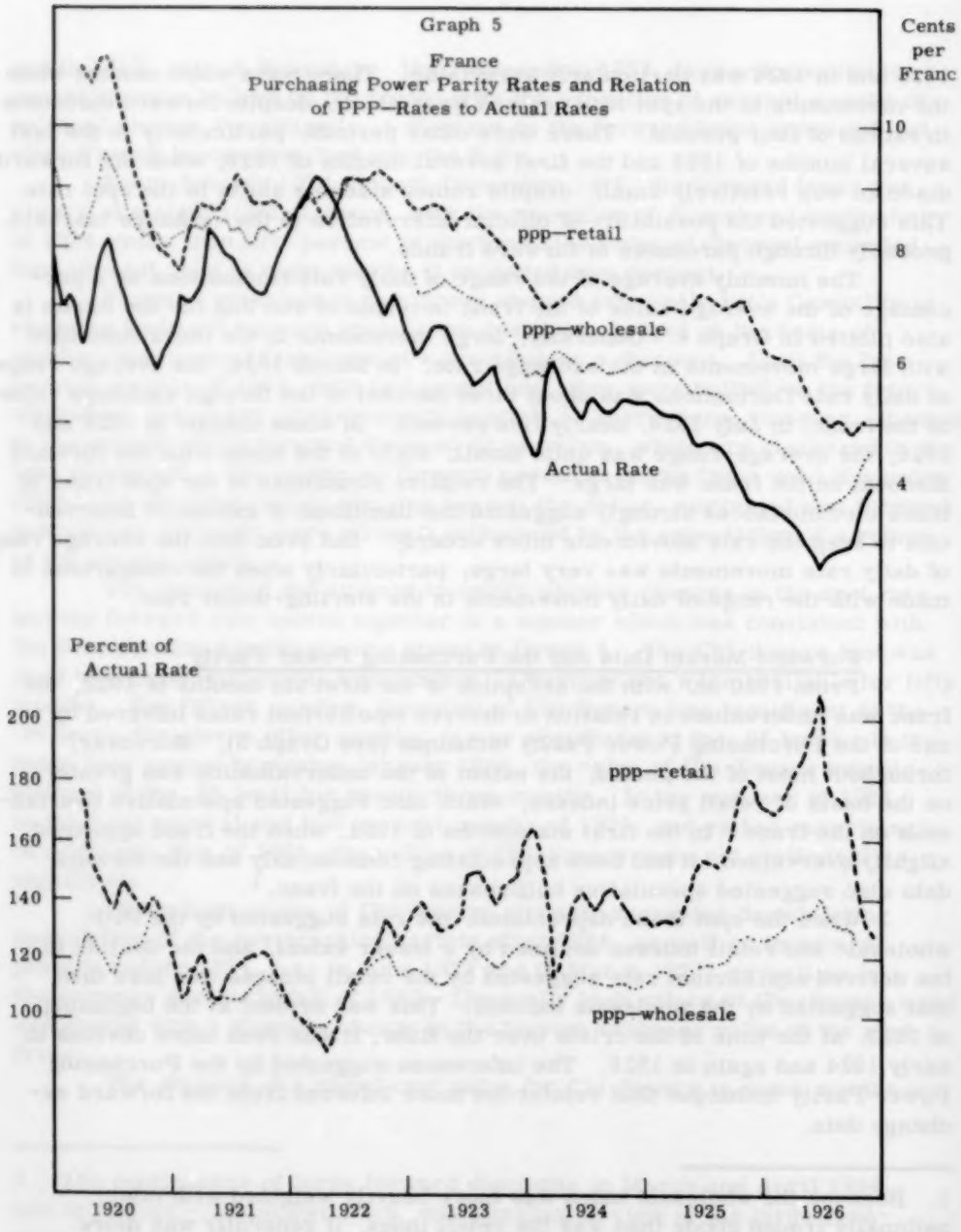
Forward Market Data and the Purchasing Power Parity

From 1920 on, with the exception of the first six months of 1922, the franc was undervalued in relation to derived equilibrium rates inferred by use of the Purchasing Power Parity technique (see Graph 5). Moreover, throughout most of the period, the extent of the undervaluation was greater on the basis of retail price indexes, which also suggested speculative bearishness on the franc. In the first six months of 1922, when the franc appeared slightly overvalued, it had been appreciating considerably and the forward data also suggested speculative bullishness on the franc. [1]

When the spot franc depreciated, the rate suggested by the PPP wholesale and retail indexes declined by a lesser extent, and the decline in the derived equilibrium rate suggested by the retail indexes was less than that suggested by the wholesale indexes. This was evident at the beginning of 1923, at the time of the crisis over the Ruhr; it was even more obvious in early 1924 and again in 1926. The inferences suggested by the Purchasing Power Parity technique thus reinforced those inferred from the forward exchange data.

---

1. Because the wholesale index was more heavily weighted with internationally traded goods than was the retail index, it generally was more greatly affected by changes in the exchange rate. The two instances when the rate suggested by the wholesale index suggested a higher rate for the franc occurred when the franc appreciated sharply, leading to a greater decline in French wholesale prices than in the retail prices, so that the PPP wholesale rate exceeded the PPP retail rate. Speculators were then bullish on the franc.


Graph 5
France
Purchasing Power Parity Rates and Relation
of PPP—Rates to Actual Rates
Cents per Franc
10
8
6
4
ppp—retail
ppp—wholesale
Actual Rate
Percent of Actual Rate
200
180
160
140
120
100
ppp—retail
ppp—wholesale
1920
1921
1922
1923
1924
1925
1926

The Impact of Speculation

The evidence from the forward market, supported by that based on the Purchasing Power Parity, strongly suggested that speculative forces in the exchange market were destabilizing of the overtracking variety. After October 1923, the bearishness of forward speculators varied directly with the fluctuations of the franc in the exchange market, and greatly contributed to these fluctuations. Speculative sales of francs after October 1923 were one of the causes of the considerable undervaluation of the franc, and the speculatively induced decline of the franc in the exchange market also resulted in upward pressures on French prices.

This technical evaluation of the character of speculation, however, was incomplete in itself. Speculators who were bearish on the franc were obviously proved "correct" by the decline in the franc. It was desirable, therefore, to determine whether the franc would have declined similarly in the absence of speculation, or whether it was the impact of speculation which proved so disastrous to its foreign exchange value.

Speculative shifts from the franc might seem justified by the depreciation of the real value of French monetary assets after June of 1922. One possible cause of the increase in the price levels within France was the continued large government deficits.[1] During the war, only about fifteen percent of the government's expenditures were tax-financed, on the presumption that borrowings would be repaid from reparations. After the Armistice a Recoverable Budget (the German budget) was established to finance war reconstruction expenditures and transfers; it, too, was loan-financed in the belief that reparations would be available soon. In 1920, forty percent of total government expenditures were in this budget.

---

1. Examination of the change in the Treasury's domestic debt outstanding, with adjustments for changes which do not reflect current expenditures, provided the best evidence for the size of the fiscal deficit or surplus. Tracing through the maze of government finance was complicated because the three major government budgets (Ordinary, Extraordinary, and Recoverable) were encumbered by twelve annex budgets, covering State Railways and the like, special Treasury budgets, and supplementary credits attached to budgets of earlier years.

Moreover, certain transactions, such as the extension of non-negotiable annuities to injured war veterans and war orphans, were treated as expenditures for the full amount of the annuity although only in subsequent years as interest was paid was there a charge on the government's cash position. Other transactions, which were effectively authorizations rather than appropriations, were treated as an expenditure in the year authorized, although the cash payment may have lagged by two, three, or even four years. See Roger Haig, The Public Finances of Postwar France (New York: Columbia University Press, 1929) Chapter XX.

By 1923 it became obvious that the yield from reparations would be very much less than anticipated. This realization, together with the recognition that the government deficits were contributing to higher prices within France, stimulated intense efforts to reduce the deficit, and it declined from about 12 billion francs in 1923 to slightly over 2 billion francs in both 1924 and 1925.[1] In these two years, about eight percent of the government's expenditures were debt-financed.

The price-raising impacts of the large government deficits prior to 1924 were less obvious than it might immediately seem; in the presence of idle productive resources the deficit may have exerted its major impact on the level of output. The French retail price index in June 1923 was at the level it had been at in December 1920, having declined until the end of 1922 and then risen during 1923. The period of large government deficits had nearly ended before retail prices began to increase significantly.

Two factors which were relevant to the price-raising impacts of the government deficits were the extent to which these deficits were bank-financed and the price-raising impact of the expenditures themselves. The evidence strongly suggested that almost all of the increase in the government debt was taken up by the public and non-bank financial institutions. At the beginning of 1922, the advances from the Bank of France to the Treasury were 24.2 billion francs, a level not reattained until May 1925. Unfortunately, data on commercial bank holdings of government debt were not available but, since the demand liabilities of the four largest banks only increased from 13.2 billion francs at the end of 1920 to about 15 billion francs at the end of 1923 and 1924, they could only have absorbed a relatively small part of the new government issues.

Much of the deficit appeared to be an offset to the desire of the public to accumulate financial assets, and the Treasury appeared not to have needed to rely on the banking system for funds to enable it to compete with the public for resources.

In the absence of adequate data on output capacity, including that part of the labor force unemployed, it was difficult to determine if the demands of the government, partly financed by borrowing, would have led to higher prices. Because of wartime destruction and the annexation of Alsace-Lorraine, the prewar base of the index of industrial production may be misleading. On the 1913 base, this index reached 100 only at the beginning of 1924. It was generally acknowledged that 1921 was a slump year, and that 1922 was the beginning of an extended recovery. If production had been close to capacity in 1922 and 1923, it might be expected that the index of building activity would have risen to its prewar levels earlier than it did (see Table 3). Moreover, the very rapid rate of increase in industrial

1. In comparison, the flow of short-term capital from France was estimated to be between 7 and 10 billion francs in 1924; between 10 and 12 billion francs in 1925; and, for the first seven months of 1926 alone, at 17 billion francs. Pierre Meynial, Créances et Dettes Internationales (Paris: Librairie Dalloz, 1926), pp. 119–154, and esp. 135.

TABLE 3

Selected Economic Statistics, France 1919–1926
(1913 = 100)

	Retail Prices	Wholesale Prices	Wholesale Prices Domestic Component	Wholesale Prices Imported Component	Industrial Production	Building Activity
1919 June	263	329	368	276	62	13
Dec.	285	423	421	452	62	22
1920 June	369	493	507	498	63	32
Dec.	324	435	485	370	67	27
1921 June	312	325	379	247	50	21
Dec.	323	326	357	288	62	26
1922 June	307	325	363	276	79	31
Dec.	305	362	380	351	88	44
1923 June	331	409	434	387	87	70
Dec.	365	459	469	467	98	57
1924 June	370	465	475	475	103	84
Dec.	404	507	494	561	113	62
1925 June	422	543	533	591	105	89
Dec.	463	632	587	752	118	84
1926 June	544	738	682	883	126	133
Dec.	599	626	647	628	123	108

Sources: International Abstract of Economic Statistics, The Process of Inflation in France.

production after 1922 suggested the presence of unused capacity. This production data, to the extent it merited any attention, suggested that the demand for resources by the government probably contributed relatively little toward the increase in French prices prior to 1924.

The years 1922 and 1923 provided an interesting comparison between the size of the government deficit, the upward movements in prices, and the change in the value of the franc in the exchange markets.[1] In 1922,

1. The impact of changes in the foreign exchange value of the franc on the domestic price level was revealed by the decline in the French wholesale price index from 600 in April 1920 to 330 in June 1921—a decline of

the net government deficit was about 15 billion francs; in 1923, about 12 billion. In 1922, French retail prices remained relatively constant; in 1923, they increased by twenty percent. In 1922, the franc depreciated by about ten percent against the dollar; in 1923, it depreciated by forty percent.

The depreciation of the franc began to have a cost-push impact in France, and the wholesale price index began to be pulled up by its import components.[1] From June to December 1922, the latter index moved from 276 to 351, about twenty-five percent, after having declined slightly in the previous six months. The wholesale index increased from 325 to 363 in the last six months of 1922, after having remained relatively unchanged over the first half of the year. For each of the following years, the import component of the wholesale index increased considerably more rapidly than the wholesale index, thus pulling it upward. The upward pressure on prices in France seemed much more a result of the cost-push impact on French prices from the increased cost of imports than from the demand-pull of the domestic sectors.

If the exchange rate had been stabilized at a realistic rate after the bear raid in 1924, and after near-balance had been achieved in the government's fiscal accounts, the second round with the exchange speculation might have been avoided. Given the exchange rate, French prices in May 1924 were not excessively high in relation to foreign prices: indeed, the franc appeared undervalued on the basis of the Purchasing Power Parity technique.

In May 1924, a national election brought to power a weak leftist coalition, the Bloc des Gauches, which never commanded an effective parliamentary majority. In the following two years the government fell six times, primarily over proposals designed to secure domestic economic stabilization. The franc continued to depreciate until Poincaré returned to power as the head of a Government of National Union in July 1926, in the midst of the spectacular collapse of the franc.

---

forty-five percent. During this period, the French government deficits were extremely large, but their price raising impact was more than offset by the lower cost of imports. Between the same dates, the U.S. wholesale price index declined by forty-four percent.

1. The French wholesale price index can be broken down into an imported goods component and a domestic goods component. Until the middle of 1923, with the exception of the last several months in 1919, the domestic component was higher (see Table 3). After the first half of 1924, however, the prices of imports in the index began to exceed the prices of the domestic goods, and this divergence increased considerably as long as the franc depreciated. The excess of the domestic index over the imported index until 1923 suggested that the deviation was caused by demand-pull factors within France; the excess of the imported component after June 1924 suggested cost-push factors coming from the exchange rate.

The two major domestic financial problems faced by the successive governments were the need to maintain a satisfactory balance in the government's fiscal accounts and the need to keep the public from cashing any significant amount of its holdings of government securities. The successive, and increasingly frequent, government crises reflected the inability of the various governments to obtain the legislative support for their proposals designed to accomplish these objectives. The anomalous situation was that, as the government's deficit became smaller, the debt management problem became more difficult. The political uncertainty resulting from the government crises was a major factor leading to the flight from the franc but, in large measure, the speculation against the franc, because of its upward impact on French prices, was an indirect cause of these crises.

The fiscal problem was one of increasing government taxes simply to maintain the near-balance in budgetary accounts in a period of rising prices. The French tax system was extremely regressive; given the tax base, the increase in tax receipts as a result of rising prices was much less than proportional to the increase in prices.[1] Even if tax receipts had increased only slightly less than proportionally to the increase in prices, the fiscal problem would have eased because between forty and fifty percent of the government's payments were for servicing its debt, and these costs were relatively fixed. While the costs of servicing the external debt increased as the franc depreciated, not much of this debt was serviced.

The uncertainty resulting from these political setbacks complicated the debt management problem, and the rising prices which resulted from the speculation against the franc and the capital outflow had an even more damaging indirect effect. During and immediately after the war, the Treasury had had some success in funding part of its debt. After the beginning of the long upward swing in prices in the summer of 1922, selling longer term issues became progressively more difficult, even though the coupon and other features brought net yields as high as eight or nine percent. With the longer-term markets virtually closed, the Treasury

---

1. This table shows the relation between actual tax receipts and what tax receipts would have been if tax receipts had increased as rapidly as the retail price index. 1922 was used as a base year. The legislated tax increases of 3.3 billion francs in 1924, an additional 4.6 billion in 1925, and 3.5 billion more in 1926 are included in the actual receipts to the extent that they were collected. The computed receipts ignored this change in the tax base.

	1922	1923	1924	1925	1926
Actual receipts	20.8	20.9	25.3	29.8	38.1
Computed receipts	20.8	23.0	26.2	29.4	38.3

Source: The Public Finances of Postwar France, pp. 91, 108, 126, and 396.

TABLE 4

Selected Monetary Statistics, France 1920–1926
(Billions of Francs)

	Note Circulation	Demand Deposits	Money Supply[1]	Floating Debt of the Treasury[2]	Advances of Bank of France to Treasury	Commercial Bills Discounted by Bank of France	Current Accounts at Bank of France	Official Discount Rates (%)	Long-Term Interest Rates[3] (%)
1920									
Dec.	37.9	13.2	50.1	47.3	26.6	3.3	3.5	6	
1921									
June	37.4	12.6	50.0	55.3	25.0	2.9	2.7	6	
Dec.	36.5	13.3	49.8	60.4	24.6	2.5	2.7	5.5	
1922									
June	36.0	13.8	49.8	65.7	23.3	2.3	2.4	5	6.3
Dec.	36.4	13.9	50.3	59.1	23.6	2.4	2.3	5	6.2
1923									
June	36.7	14.2	50.9	58.9	23.1	2.4	2.1	5	6.2
Oct.	37.7	14.4	52.1	58.8	23.4	3.1	1.9	5	6.8
Nov.	37.3	15.2	52.5	57.4	22.8	3.7	2.2	5	6.7
Dec.	37.9	14.9	52.8	57.7	23.3	3.7	2.4	5	6.6
1924									
Jan.	38.8	15.1	53.9	58.4	22.8	4.2	2.3	6	6.4
Feb.	39.3	15.5	54.8	58.6	23.1	3.9	2.3	6	7.0
Mar.	39.9	15.5	55.4	58.3	22.7	5.4	3.2	6	7.5
Apr.	39.8	15.2	55.0	59.5	22.7	4.6	2.4	6	7.3
June	39.7	15.4	54.8	60.2	23.0	3.7	2.1	6	7.4
Aug.	40.0	16.1	56.1	61.7	22.8	4.4	1.9	6	7.5
Oct.	40.5	15.3	55.8	58.4	22.7	5.5	1.9	6	7.5
Dec.	40.6	14.8	55.4	56.4	22.6	5.2	1.9	7	8.1

Dec.	40.6	14.8	55.4						
1925									
Feb.	40.8	15.1	55.9	56.8	21.9	5.6	1.9	7	8.1
Apr.	43.0	15.3	58.3	54.9	23.2	5.5	1.9	7	8.4
June	43.0	15.9	58.9	53.8	25.7	3.7	2.1	7	8.2
Aug.	44.7	17.7	62.4	54.8	27.8	3.2	2.4	6	8.1
Oct.	46.7	19.0	65.7	50.3[4]	29.9	2.9	2.2	6	8.9
Dec.	51.1	19.9	71.0	48.1	35.9	3.4	3.2	6	8.3
1926									
Feb.	50.9	20.9	71.8	48.8	34.5	3.1	2.8	6	10.3
Apr.	52.2	20.4	72.6	48.9	35.2	3.5	2.7	6	9.8
May	52.7	21.4	74.1	48.9	35.9	4.5	2.8	6	8.9
June	53.1	22.4	75.5	48.8	36.6	4.5	2.8	7.5	9.8
July	56.0	23.3	79.3	46.7	37.5	5.2	3.2	7.5	10.1
Aug.	55.1	22.7	77.8	48.0	36.5	5.9	3.4	7.5	-
Oct.	54.5	23.1	77.6	46.8	35.8	5.0	2.9	7.5	9.8
Dec.	52.9	22.5	75.4	49.5	36.0	3.6	5.2	6.5	9.2

1. This is the sum of the first and second columns.

2. Primarily Bons de la Defense Nationale.

3. New issues.

4. About this time 6 billion of Bons de la Defense Nationale were exchanged for a Rente whose coupon was pegged to the sterling-franc exchange rate.

Sources: International Abstract of Economic Statistics, The Public Finances of Postwar France.

TABLE 5

The Financial Position of the French Treasury 1919-1926

(Millions of Francs)

	1919	1920	1921	1922	1923	1924	1925	1926
Fiscal Surplus (Deficit (-))	-48,655	-37,686	-21,360	-15,315	-11,962	-2,125	-2,213	3,193
Foreign Sources (Uses (-))	9,672	-1,323	-1,514	-1,238	-847	-985	-659	-1,150
Domestic Sources (Net)	38,983	39,009	22,874	16,553	12,809	3,110	2,872	-2,043
Source of Domestic Funds:								
Perpetual Securities		11,278					6,011	
Intermediate and Long-Term Securities	4,613	10,915	1,871	7,822	4,951	1,567		
Short-Term Securities	24,454	15,569	22,724	8,911	7,431	3,294		240
Public Deposits at Treasury	1,546	617	379	670	1,031	532	1,218	495
Advances from Banks	8,370	865					14,150	1,050
Total New Sources	38,983	39,244	24,974	17,403	13,413	5,393	21,379	1,785
Uses of Domestic Funds:								
Retirement of Bank Advances		235	2,100	850	604	592	634	668
Retirement of Intermediate Debt						1,691	8,064	3,160
Retirement of Short-Term Debt							9,809	
Total New Uses		235	2,100	850	604	2,283	18,507	3,828

Note: Certain figures in this table are inconsistent with those in Table 4. Since the interpretation does not depend on the resolution of these inconsistencies, the matter has not been pursued.

Source: The Public Finances of Postwar France

was forced to rely on subscriptions to its short-term issues as the only alternative to a sale of debt to the banking system.

In 1925, about 21 billion francs of short-term bonds matured, and in 1926 an additional 3 billion came due.[1] Moreover, the floating debt, primarily the Bons de la Defense Nationale, also had to be continuously refinanced. The Bons were negotiable Treasury Bills of one to twelve months' maturity, available in a wide range of denominations. They were particularly designed for small savers and could even be discounted prior to their maturity at a very small cost. Nearly 58 billion francs of Bons were outstanding at the end of 1923. The average maturity on the floating debt was extremely short, and the Treasury had to sell about 160 billion of the floating debt annually to prevent the amount outstanding from declining.

Because any attempt to tighten credit would simply have induced commercial lenders and borrowers to reduce their holdings of government securities, forcing the Treasury to borrow from the Bank of France, the Bank was precluded from having an effective monetary policy.[2] The Bank remained the emergency source of finance for the Treasury, although legislation limited both its advances to the Treasury and its own note issues, and also provided for fixed annual decreases in these limits. The potential demand on the Treasury was extremely large, and yet the extent of its recourse to the Bank of France was extremely limited.

As long as the public's money holdings were excessive, they were able to finance a larger volume of money transactions without having to dispose of their Bons. However, as their ability to finance a larger volume of money transactions with the existing money supply diminished, more Treasury securities were redeemed for cash.

The management of the debt did not become a major problem for the Treasury in the first exchange crisis which ended in March 1924. Even though the franc lost nearly one-half its value in about five months, the decline was so rapid and abrupt that prices failed to keep pace; wholesale prices rose by less than thirty percent and retail prices by less than fifteen percent over the same period. Moreover, the rise in prices was much greater than the increase in the money supply, which only went up by about six percent. There was only a negligible decline in the public's holdings of Bons.[3] Because of the lag between the movements in the exchange rate

---

1. Haig, op. cit., p. 107.

2. However, the Bank of France, unlike the Bank of England, had no strong tradition of using its discount rate to protect its reserves. Prior to World War I, it increased its gold selling price rather than its discount rate whenever it wished to discourage a gold drain; it was commonly said to be on the "limping gold standard." Commission on Gold and Silver Inquiry, European Currency and Finance, Washington, 1925, p. 308.

3. Dulles, op. cit., p. 186.

and the movements in the price level, and because of the public's large holdings of liquid assets, the exchange rate could fall nearly fifty percent, lead to a percentage increase in the money supply one-tenth as large, and still result in only limited upward pressure on interest rates.

Refinancing the Treasury's debt after the end of 1924 became much more difficult as a result of the increased need of the public for money balances as prices rose and business activity increased. The public and the banks were able to increase their cash holdings simply by not resubscribing when some of their holdings of Bons and short-term debt matured, forcing the Treasury to go to the Bank of France.

The Treasury was effectively cornered between the limitations on the advances of the Bank to the State and on the note circulation of the Bank, and the increased need for money balances on the part of the public. In the first several months of 1925, the Treasury attempted a maneuver which previous governments had used as a means of circumventing the limitation on the advances of the Bank of France to the Treasury. The commercial banks were induced to increase their commercial discounts at the Bank, which freed funds to enable them to buy more governments. As this practice continued, the note circulation began to press against its limit. In order to get the cash necessary to pay its current bills and its maturing debt redeemed for cash, the Treasury induced the Bank of France to falsify its balance sheet by understating its note issue. When this strategem became known in April 1925, the government fell, and the future of the franc became all the more suspect.

To regain control over the money supply, it was necessary to break the chain linking speculation against the franc and rising prices within France, increased government expenditures, and redemption of the government debt. In 1925, various governments attempted to secure budgetary balance through increases in taxes and through governmental economies, and were at least partially successful. While attainment of near-balance in the fiscal accounts was a necessary condition for stabilization, it was not sufficient.

Various attempts were made to break the links by attacking the floating debt directly. A capital levy was mentioned continually, as was forced consolidation of the short-term debt. Even the mention of these proposals stimulated a further flight from the franc.[1] In the late summer of 1925, the government offered a form of purchasing power bond, with a

---

1. The external value of the franc was greatly influenced by the man serving as Premier, or designated to form a government, primarily because of his views on these controversial proposals. Herriot, for example, was a strong advocate of the capital levy; when he was called to power in July 1926, the franc fell by nearly twenty-five percent in five days. When his government fell, the franc climbed just as rapidly, even before a new cabinet had been formed. Dulles, op. cit., pp. 194-195.

coupon tied to the sterling exchange rate. It was expected that perhaps 30 billion of the outstanding Bons might be offered in exchange for this variable yield security, but only 6 billion was actually subscribed and the offer was deemed a failure. In the climate of the day, the French security holder was extremely suspicious of any long-term commitment to government securities. The short-term debt, however, was not the cause of the problem, but rather the symptom. The achievement of price stability might have provided a solution for the debt management problem; this would have entailed either reducing domestic demands or preventing further depreciation of the franc in the exchange market.

Intervention in the exchange market in 1925 and 1926 was another attempt to break the links. While such intervention may have had the effect of delaying the rate of depreciation of the franc, the volume of resources committed and the boldness of the intervention were not adequate to stop or reverse the capital outflow. To stop any further increases in domestic prices, intervention would have had to have been sufficiently bold to achieve an appreciation of the franc, for this was necessary to minimize or offset the lagged upward pressure on prices resulting from the previous depreciation of the franc.

July 1926 was a month of crises, and the franc fell as low as 240 against sterling. Poincaré eventually was asked to head the third government in a month and came to power as head of a Government of National Union, embracing a wide range of parties. He thus secured support for his proposals for vigorous financial reform. This resulted in a gain of confidence and induced a large capital inflow. Within slightly more than four months the franc had climbed to 125 per pound, despite the slowing of the rate of appreciation from large official purchases of foreign exchange. Eventually, the franc was stabilized at 122 against sterling, and at this level it was generally regarded as undervalued.

Conclusion

The data examined—the level of the forward rate, the relation between the movement in the spot rate and the movement in the forward rate, the deviations of the actual rate from an inferred equilibrium rate derived by using the Purchasing Power Parity technique, and the average daily range of movement in the spot rate—all suggested disruptive speculation in the French franc after October 1923. Moreover, even before forward speculators became bearish on the franc in late 1923, the average daily range of movement in the spot rate was exceedingly large, and the franc already undervalued. The PPP data also suggested that speculative transactions had increased the amplitude of cyclical fluctuations in the rate.

After forward speculators became bearish on the franc in late 1923, it came under severe speculative pressures. Within five months it had lost nearly one-half of its value in terms of the dollar. Thereafter, this speculatively induced increase in the price of foreign exchange became the largest source of upward pressure on French prices. In late 1923 and early 1924, however, the increase in the price indexes greatly lagged the increase in the

price of foreign exchange. Because the public already had large cash holdings, relatively little restrictive monetary pressure resulted from the increased need for cash balances. The prospective speculative gain from selling francs apparently greatly exceeded that from holding or buying various domestic assets.

Even if the money supply had been absolutely fixed, the situation would not have been greatly different. The reduction in the public's liquidity and the upward pressure on interest rates, which occurred as prices rose in 1925 and 1926, would have occurred somewhat earlier. Given the speed with which the exchange rate changed and the lag between the movement in the exchange rate and the movement in the price indexes, it seems unlikely that a somewhat more rapid increase in domestic interest rates would have affected the choice between acquiring foreign assets and holding or acquiring domestic assets so as to diminish greatly the rate at which the price of foreign exchange increased.

The sudden, large change in the exchange rate was partly attributable to the reversal of the position of forward speculators from being bullish on the franc to becoming bearish on it. Because the exchange market was already an exceedingly thin market, virtually all the support given the franc appears to have come from those taking advantage of the price differentials between French goods and foreign goods resulting from the change in the exchange rate. Because reaction to such differentials was slow, the franc became extremely undervalued.

The direct impact of the weak governmental financial policies toward the fall of the franc was suffusing the public with liquidity. The indirect impact, because of the association between large government deficits, increases in the note issue, and increases in the domestic price level, was to reduce the consensus about the future stability of the franc both internally and in the exchange markets. In a relatively thin exchange market, the franc was very susceptible to a speculative attack.

With the deficit in the government's budget largely eliminated, the only justification for further speculation against the franc—since it was already greatly undervalued—was a belief that speculative forces would continue to dominate the exchange rate movement because government policies were inadequate to cope with them. From late 1924 on, the weakness of government policy was its inability to break the chain of an increase in the price of foreign exchange, higher domestic prices, increased government payments and a larger government deficit, and redemption of publicly owned short-term government debt. Only by forestalling the speculatively induced increase in the price of foreign exchange could this chain have been broken. The weakness of government financial policy was its inability to deal effectively with speculation. Yet, there can be little doubt that the major cause of higher prices, of the increase in the money supply, and of the recurrent fiscal and debt management crises was the impact of exchange speculation on the domestic price level. The impact on French prices of a much higher cost of imports, from the speculatively induced increase in the cost of

foreign exchange, exceeded the price-raising impact of the export surplus generated to finance the capital outflow; the inflationary impact of this export surplus, in turn, greatly exceeded the impact of the government fiscal deficit.

The profitability of bear speculation in the franc, facilitated by the weakness of the French financial policies, does not prove the accuracy of speculative foresight. Instead, it demonstrated that exchange speculation can provide its own justification. While the government budgets after 1923 might have resulted in some upward pressure on French prices, speculative transactions did much more than simply hasten the adjustment to a new equilibrium rate; they greatly affected the level of the equilibrium rate.

6. The Belgian Franc, 1919–1926

Belgium was a member of the Latin Union prior to World War I. The dollar parity of the Belgian franc, 19.29 cents per franc, equaled that of the French franc. Almost immediately upon the outbreak of the war, Belgium was occupied by the Germans, and the Belgian franc was no longer quoted in the exchange markets. For a short period after the end of the war, a new Belgian franc was supported at about the same level as the French franc. Official support for the Belgian franc terminated on March 1919 and was not resumed until September 1925, when the Belgian authorities attempted to stabilize the franc at 4.55 cents per franc (107 francs per pound). This support terminated in March 1926, and the franc depreciated rapidly to a low of 2 cents in July 1926. It began to appreciate with the French franc after the end of the July confidence crises in Paris, and it was eventually stabilized at 2.78 cents per franc (175 francs per pound) in October of 1926.

The movements of the Belgian franc were very similar to those of the French franc. Until 1922, they were seldom apart by more than five percent. The Belgian franc experienced the same major swings as the French franc; it depreciated rapidly in the winter of 1923–24, fell precipitously in the summer of 1926, and moved together with the French franc in many of the smaller cyclical movements. Beginning in 1921, the French franc moved to a premium in relation to the Belgian franc, and with the exception of the period when the Belgian franc was temporarily stabilized, it remained at a premium through 1926.

The causes of the solidarity of the two francs were the subject of much concern at that time, and the belief that the weakness of the French franc pulled the Belgian franc along with it was one of the major motives for the attempt to stabilize the Belgian franc in September of 1925. At the time, the solidarity of the two francs was believed to be a result "of a part of a chain of many interrelated economic and political factors," primarily involving large deficit financing for post-war reconstruction in the expectation that the Germans would eventually pay. The belated realization that they would not led to considerable suspicion of the stability of the value of both currencies.

The financial situation in Belgium had some of the same characteristics as in France. There were large government deficits prior to 1924 and small ones thereafter. The Belgian tax system was also regressive.[1]

The similarity of the movements of the two francs should not detract from several differences. Because of the occupation, Belgium emerged from the war less scathed physically than France,[2] and the Belgian government was virtually precluded from saddling itself with a large internal debt during the war. However, at the end of the war, it converted the wartime issues of occupation marks in Belgium into francs. This currency conversion was primarily financed by Treasury borrowing from the National Bank. This issue of new francs exceeded the demands for cash balances, and so a large amount went back to the Treasury as individuals bought Treasury securities and time claims on the various financial institutions, which also invested in Government securities. Except for this instance, the Belgian Treasury was able to finance its postwar deficits until 1926 without further borrowing from the National Bank. Even then, its resort to the National Bank was primarily to repay maturing debt rather than to finance its budgetary deficit.

The Exchange Market Data

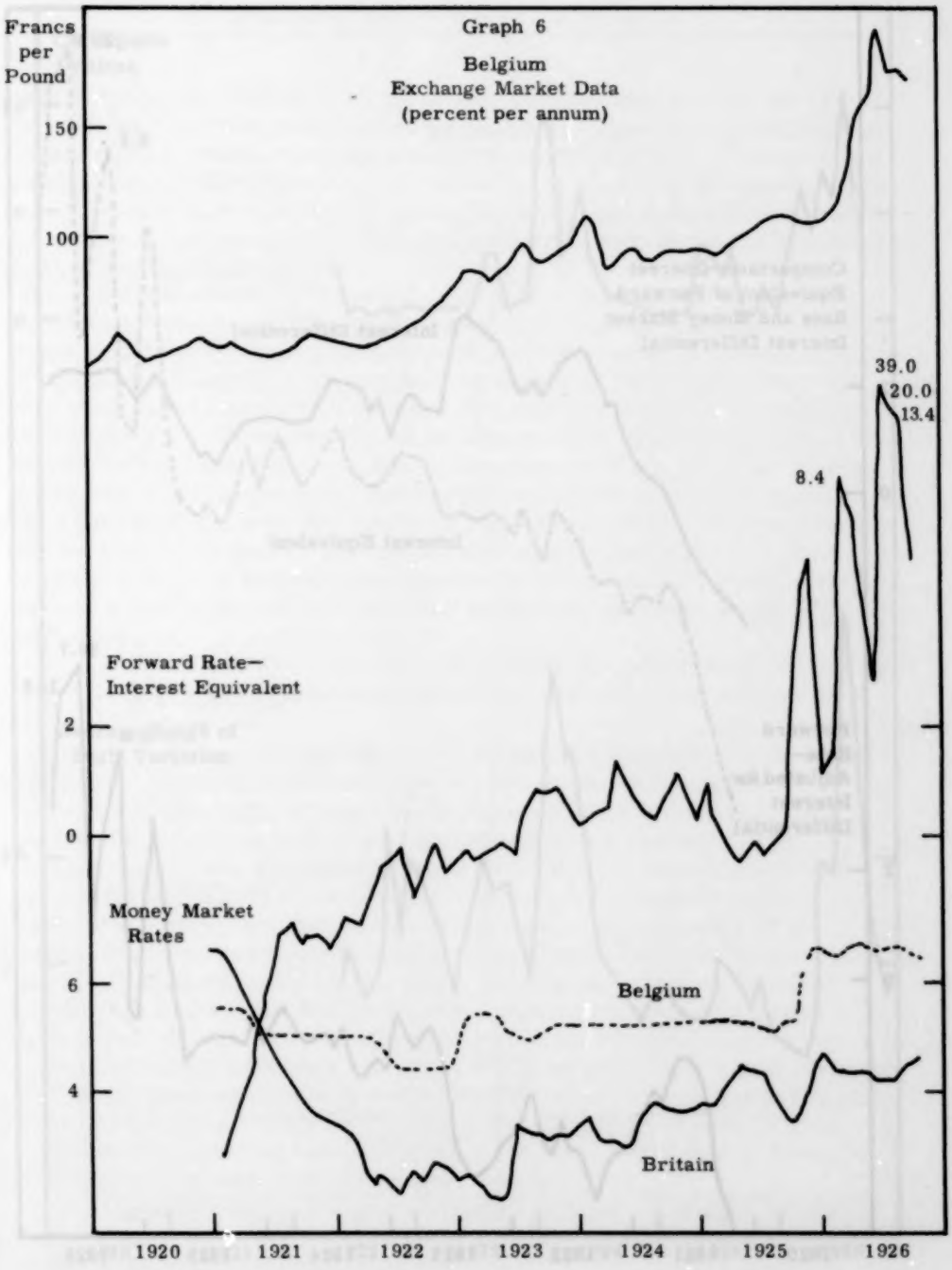
The primary foreign market for the Belgian franc was in Paris, and many transactions involving Belgian francs and other currencies involved an intermediate transaction in French francs. Purchases of Belgian francs with sterling or dollars were likely to take the form of a purchase of French francs with the sterling or dollars, and a purchase of Belgian francs with the French francs. The London price of the Belgian franc was thus really the sterling equivalent of its price in Paris, plus a small charge for the additional brokerage.

The forward market data suggested that forward speculators were extremely bearish on the Belgian franc for several weeks in March 1924, while the French franc was under attack in December 1925, and then after the failure of the attempt to stabilize the Belgian franc in March 1926 (see Graph 6). Only in this last instance was the forward franc at a discount for an extended period.

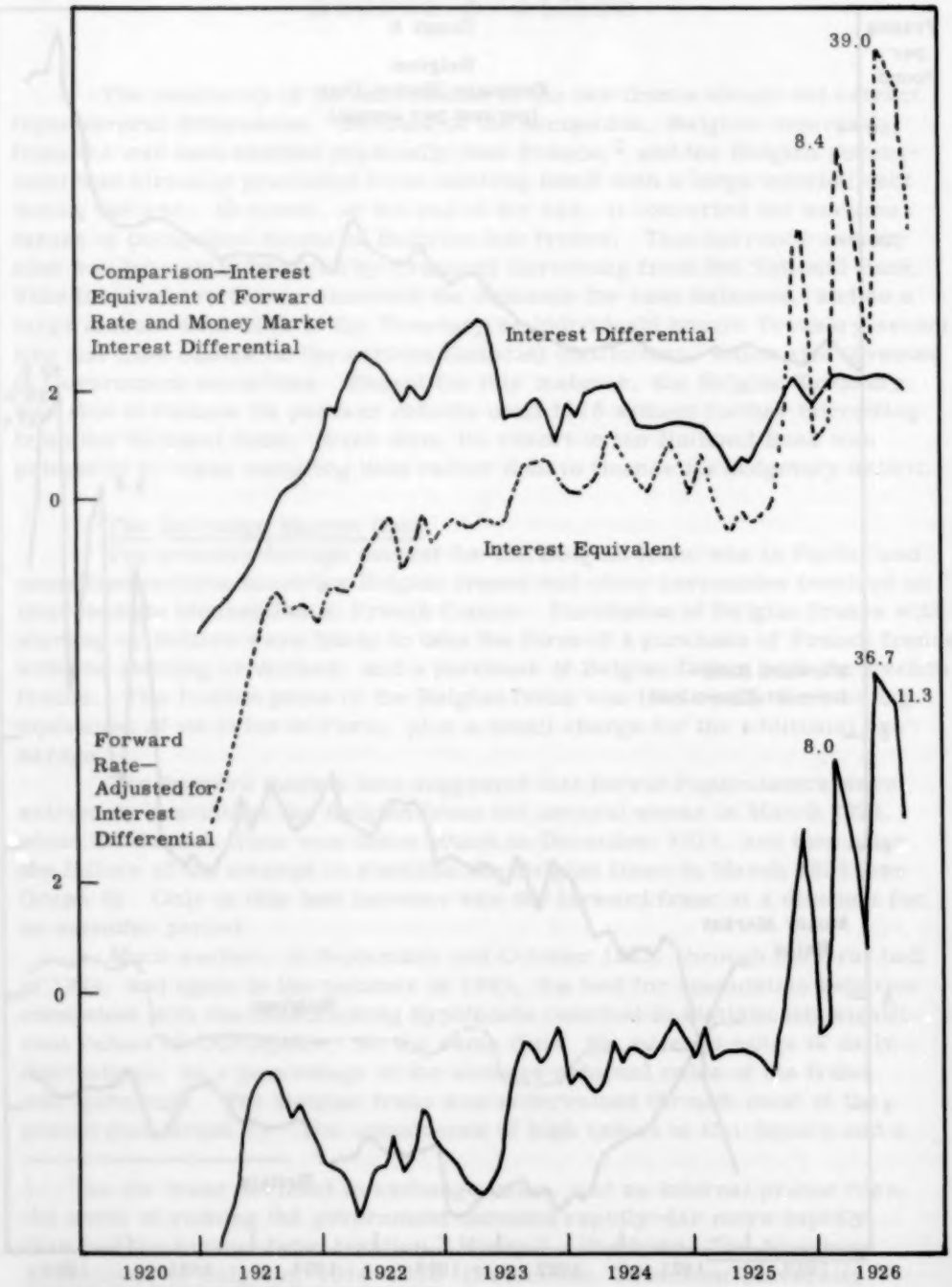
Much earlier, in September and October 1923, through the first half of 1924, and again in the summer of 1925, the test for speculative behavior consistent with the overtracking hypothesis resulted in statistically significant values of Chi-Square. At the same time, the average range of daily fluctuations, as a percentage of the average external value of the franc, was quite high. The Belgian franc was undervalued through most of the period (see Graph 7). The coincidence of high values of Chi-Square and a

1. "As the franc declined in exchange value, and as internal prices rose, the costs of running the government mounted rapidly—far more rapidly than had the income from taxation." Henry L. Shepherd, The Monetary Experience of Belgium, 1914-1936 (Princeton: Princeton University Press, 1936), p. 31.

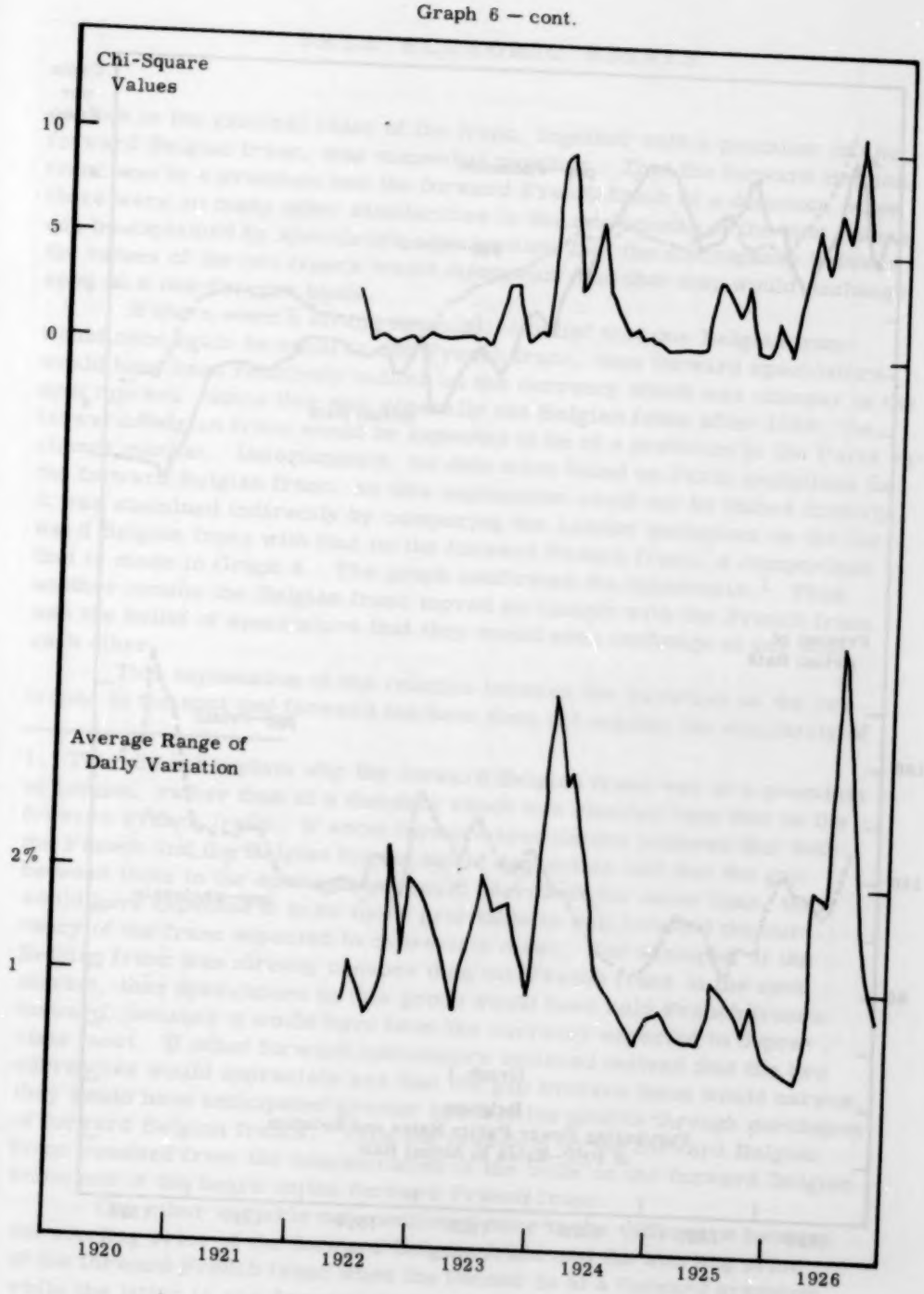
2. In 1920, the index of industrial production in Belgium reached 88; in France, 62; and 60 in Western Europe (France, Belgium, Italy, and Germany). Ibid., p. 251.

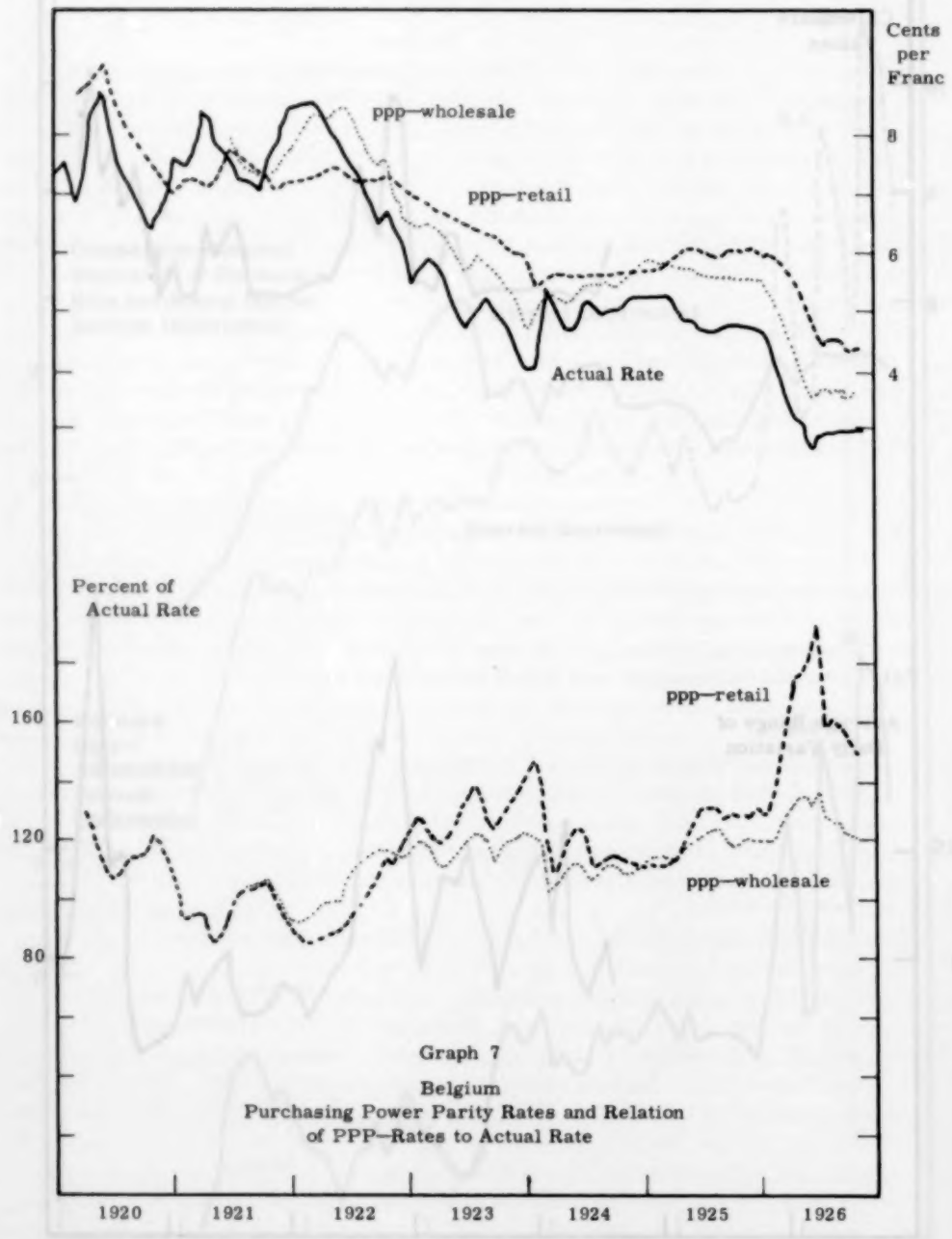

Francs per Pound
Graph 6
Belgium
Exchange Market Data
(percent per annum)
150
100
39.0
20.0
13.4
8.4
Forward Rate—
Interest Equivalent
2
0
Money Market
Rates
6
4
Belgium
Britain
1920
1921
1922
1923
1924
1925
1926

Graph 6 — cont.


39.0
8.4
Comparison—Interest Equivalent of Forward Rate and Money Market Interest Differential
Interest Differential
2
0
Interest Equivalent
36.7
11.3
Forward Rate— Adjusted for Interest Differential
8.0
2
0
1920
1921
1922
1923
1924
1925
1926

Graph 6 — cont.


Chi-Square
Values
10
5
0
Average Range of
Daily Variation
2%
1
1920
1921
1922
1923
1924
1925
1926



Graph 7

Belgium

Purchasing Power Parity Rates and Relation of PPP—Rates to Actual Rate

decline in the external value of the franc, together with a premium on the forward Belgian franc, was somewhat puzzling. That the forward Belgian franc was at a premium and the forward French franc at a discount, when there were so many other similarities in the movements of the spot francs, can be explained by speculative anticipations that the discrepancy between the values of the two francs would disappear, and that they would exchange soon on a one-for-one basis.

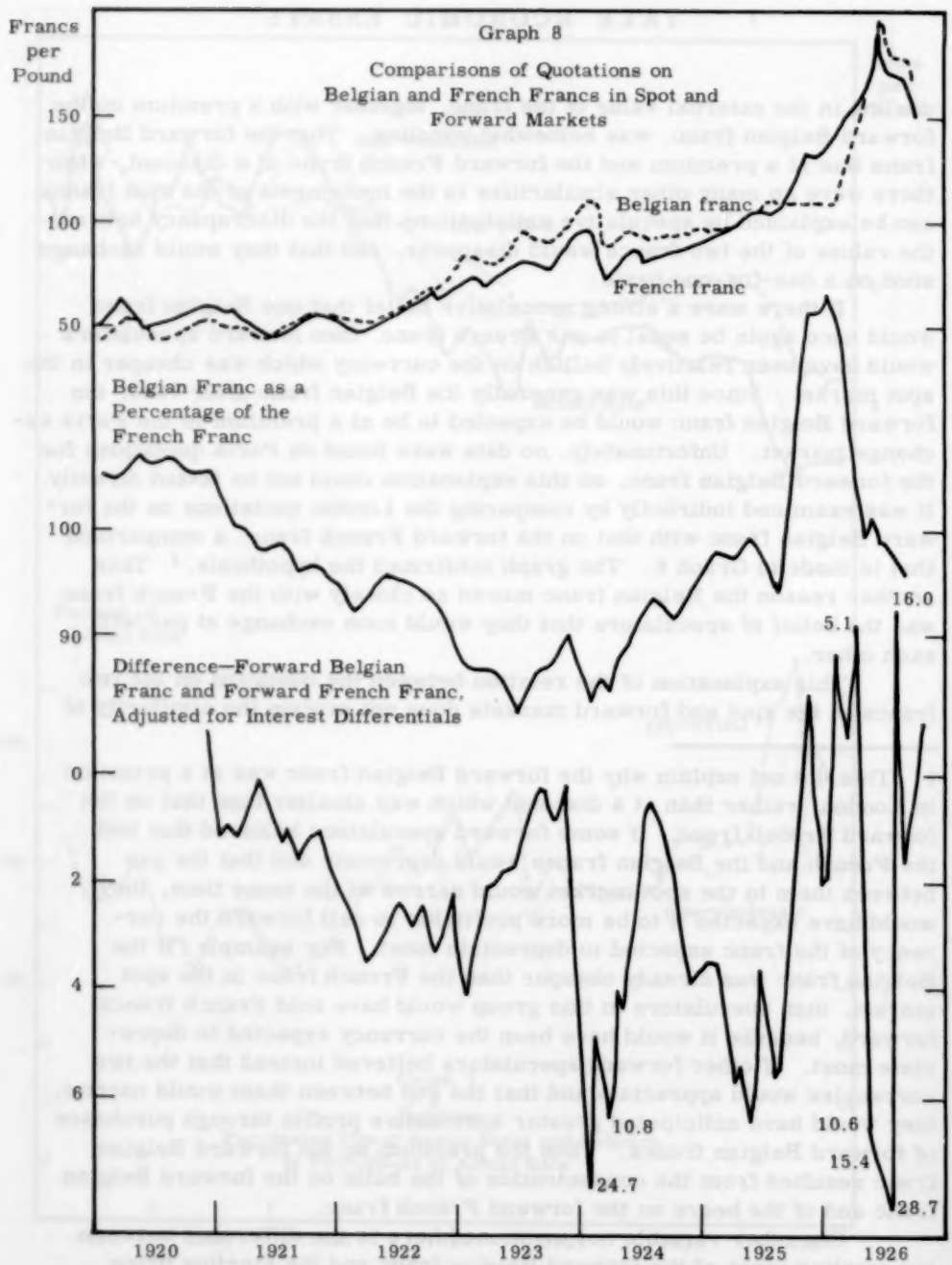
If there were a strong speculative belief that one Belgian franc would once again be equal to one French franc, then forward speculators would have been relatively bullish on the currency which was cheaper in the spot market. Since this was generally the Belgian franc after 1920, the forward Belgian franc would be expected to be at a premium in the Paris exchange market. Unfortunately, no data were found on Paris quotations for the forward Belgian franc, so this explanation could not be tested directly. It was examined indirectly by comparing the London quotations on the forward Belgian franc with that on the forward French franc, a comparison that is made in Graph 8. The graph confirmed the hypothesis.[1] Thus another reason the Belgian franc moved so closely with the French franc was the belief of speculators that they would soon exchange at par with each other.

This explanation of the relation between the quotation on the two francs in the spot and forward markets does not explain the similarity of

---

1. This did not explain why the forward Belgian franc was at a premium in London, rather than at a discount which was smaller than that on the forward French franc. If some forward speculators believed that both the French and the Belgian francs would depreciate and that the gap between them in the spot market would narrow at the same time, they would have expected it to be more profitable to sell forward the currency of the franc expected to depreciate most. For example, if the Belgian franc was already cheaper than the French franc in the spot market, then speculators in this group would have sold French francs forward, because it would have been the currency expected to depreciate most. If other forward speculators believed instead that the two currencies would appreciate and that the gap between them would narrow, they would have anticipated greater speculative profits through purchases of forward Belgian francs. Thus the premium on the forward Belgian franc resulted from the concentration of the bulls on the forward Belgian franc and of the bears on the forward French franc.

One other variable not mentioned here is the difference between the sterling price of the forward Belgian franc and the sterling price of the forward French franc when the former is at a forward premium while the latter is at a forward discount. It has been assumed that the former would be cheaper simply because the spot Belgian franc was cheaper, and this was likely to be true if the disparity between the two spot rates was more than several percent.


Francs per Pound
Graph 8
Comparisons of Quotations on Belgian and French Francs in Spot and Forward Markets
150
100
50
Belgian franc
French franc
Belgian Franc as a Percentage of the French Franc
100
90
16.0
5.1
Difference—Forward Belgian Franc and Forward French Franc, Adjusted for Interest Differentials
0
2
4
6
10.8
24.7
10.6
15.4
28.7
1920
1921
1922
1923
1924
1925
1926

their movements in the spot market, nor, more importantly, why prices in Belgium rose more rapidly than those in France, even though the financial position of the Belgian Treasury appears to have been somewhat sounder than that of the French Treasury.[1] The answers to these questions lay in the close economic ties between the two economies. Belgian and French exports were highly competitive, and Belgium tended to lose export sales whenever speculative pressures directed toward the French franc made French goods relatively cheaper in foreign markets. Belgium was a large importer of French goods, and any decline in the price of the French franc in Brussels would have tended to lead to increased imports from France. While this might explain why the Belgian franc moved with the French franc, it does not suggest why prices rose more rapidly in Belgium.

The Belgian economy was much more of an open economy than the French economy, with a foreign trade sector larger in relation to its domestic economy than was the case in France. Similar percentage increases in the price of imports into both France and Belgium would almost certainly have had a much larger cost-push impact on the Belgian economy because its domestic output embodied a larger proportion of imports than in France. Thus the same percentage change in the price of foreign exchange would have resulted in greater upward pressure on Belgian prices than on French prices. The greater increase in the prices of Belgian goods would have resulted in a somewhat lower external value for the Belgian franc than the French franc.[2]

The attempt to stabilize the Belgian franc in the autumn of 1925 was prompted by a desire to break the nexus with the French franc. However, it is unlikely that it could have succeeded, given the strong speculative belief in the solidarity of the two francs, as long as the French franc remained under speculative attack. Once the French franc depreciated below the Belgian franc, the Belgian franc came under strong speculative pressure. In effect, Belgian reserves were being committed to the support of the French franc. When these reserves were exhausted and the support of the Belgian franc withdrawn, it depreciated by fifty percent within ten weeks. Yet, Belgian prices, both the wholesale index and the retail index, had

---

1. The evidence for this view was suggestive rather than conclusive. It consisted of the lower ratio of the deficit to total government expenditures in Belgium than in France and the greater success the Belgians had selling longer term debt. Over the postwar period, seventy-five percent of the Belgian government's new issues were intermediate or long-term; during each of the years 1924, 1925, and 1926, the increase in the funded issues exceeded the increase in the internal debt. Shepherd, Ibid., pp. 36–44.

2. This explanation seemed intuitively plausible and it was verified for most of the period, with the exception of the first nine months of 1925. Between the end of 1924 and October 1925, the Belgian franc depreciated by twelve percent, but the Belgian price indexes remained virtually stable.

remained stable during the previous six months while the Belgian franc was supported at 107 francs per pound.

Near-balance in the government's financial accounts and a large measure of price stability were believed necessary forerunners to external stabilization.[1] The de facto stabilization of September 1925 was an attempt to capitalize on the success in achieving these objectives and to break the tie with the French franc. About $90 million realized from an earlier foreign loan was committed to the support of the franc, and the government sought additional foreign credits of $150 million. It also sought, and obtained, an increase in domestic taxes of twelve percent, but the inability to obtain additional foreign credits necessitated a withdrawal of official support when most of the initial $100 million had been used to finance a capital outflow.

In part, the capital outflow reflected the view that it would be impossible to stabilize the Belgian franc successfully while the French franc was still depreciating, and, in part, it reflected domestic politics. In April 1925, a weak Socialist coalition government whose electoral strength was with labor and farm groups replaced a conservative government whose interests were those of the industrial, commercial, and banking groups. In the absence of support from Belgian bankers, it was virtually impossible for the Socialist government to gain acceptable terms from foreign private bankers. Moreover, Belgian bankers apparently shifted some of their own funds abroad in an attempt to weaken the chance of success of the stabilization program.[2]

The shift from Belgian francs to foreign currencies did not result in any sizeable reduction in the money supply within Belgium. Large volumes of short-term Treasury securities were redeemed on maturity. The Treasury was able to keep its debt to the Bank relatively unchanged; the francs obtained from the sale of foreign exchange were used to repay those who did not wish to exchange their maturing securities for new issues. The reduction in liquidity resulted in an increase in interest rates; the prime rate went from about five percent in September 1925 to over seven percent in March 1926. The discount rate at the National Bank was raised twice to make the increase in the private rates effective. The potential dilemma of the authorities of whether to adopt further restrictive measures while domestic prices were stable and unemployment was rising, was avoided by the collapse of the

---

1. The increase in the retail price index from the end of 1924 to September 1925 was only two percent; in 1924, this index increased by twelve percent. Wholesale prices rose by less than five percent during 1924. The greater part of the increase in the retail index during 1924 seemed to have been a lagged reaction to the increase in wholesale prices during 1923.

2. "It is nevertheless true that the hostility of leading business men (among whom were some bankers) and of the 'bourgeoisie' to the 'democratic ministry' was one of the reasons for the failure of stabilization." B. S. Chlepner, Belgian Banking and Banking Theory (Washington: The Brookings Institution, 1943), p. 50.

stabilization effort. Conceivably, if enough reserves had been available and committed, and liquidity further reduced by the capital outflow, the continued rise in interest rates would have led to a reversal of speculative anticipations. But if the French franc had continued to fall, the Belgian economy would have become an unwilling victim of a speculatively induced competitive exchange depreciation.

During the period of the support operation, the French franc had depreciated below the Belgian franc; it was quoted at 135 against sterling as support for the Belgian franc at 107 terminated. The forward Belgian franc went to a large discount immediately, and these speculative pressures so depressed the Belgian franc that within six weeks it had depreciated to the level of the French franc. Thereafter, Belgian francs depreciated more rapidly and Belgian prices rose even higher than those in France.

Summary

The dilemma of the Belgian authorities was their apparent inability to maintain a flexible rate without having the Belgian franc move with the French franc and their inability to maintain a stable rate. In the 1925–26 period, perhaps the combination of a more conservative government with the support of the bankers and larger foreign reserves would have contributed toward the success of the stabilization. But, even under these circumstances, the maintenance of a fixed value for the Belgian franc would have been difficult and costly. As long as the French franc continued to depreciate, and became increasingly undervalued, Belgium's current account probably would have deteriorated because of the close economic ties between France and Belgium. More restrictive financial measures within Belgium, which might have been adopted to limit the drain on the reserves, would have become costly in terms of domestic economic objectives.

The Belgian experience provides an interesting example of the difficulty of maintaining economic independence in a system of flexible exchange rates. It was Belgium's unique disadvantage to be so closely involved with France that it could not pursue an independent policy. While the structural factors were one cause of this close involvement, they were greatly accentuated by the speculative factors.

7. The Dutch Guilder, 1919–1926

Because of its central location and its neutrality, the Netherlands received a large amount of foreign capital during the war; within four years the gold holdings of the Bank of Netherlands quadrupled. With a very large demand for guilders and with the increased difficulties and costs of gold shipments to the Netherlands, the guilder stayed above its prewar dollar parity of 40.2 cents per guilder for most of the war.[1] As the guilder

1. Information on this section and on section 8 was obtained from D. T. Jack, The Restoration of European Currencies (London: P. S. King & Son Ltd., 1927), Chapters IV and V, from European Currency and Finance, and from International Abstract of Economic Statistics, 1919–1930.

appreciated relative to the dollar, the incentive to ship gold to the Netherlands increased and the Dutch authorities occasionally purchased foreign currencies to prevent the guilder from appreciating excessively. The Dutch authorities felt no commitment to support the guilder at its prewar parity as it depreciated from a peak value of 53 cents in August 1918, and fell to a low of about 30 cents in November 1920, in sympathy with the decline in other European currencies. After fluctuating extensively in 1921, the guilder appreciated to its prewar parity before the end of 1922. It became weaker at the end of 1923, but before the end of 1924 it was again at the prewar parity, and hovered around it before being officially stabilized in April 1925.

While the Dutch authorities did not support the guilder at a particular level, they intervened to moderate the extent of exchange rate movements. Failing to intervene in the exchange market might have permitted large fluctuations in the guilder as foreign-owned funds were shifted abroad. Keeping the movement in the guilder orderly was complicated by doubts as to whether it should have been pegged to sterling or to the dollar. For the most part, the guilder moved with sterling, but to a lesser extent. The extent of official intervention was difficult to estimate, but the examination of the range of daily fluctuations in the sterling-guilder rate showed them to be quite small relative to fluctuations in the sterling-dollar rate. This suggested that there was extensive official intervention on a daily basis, or that private interests felt sufficiently confident of the probability of large-scale intervention to keep rate movements small to do so on their own.

The Netherlands shared many of the fiscal problems of the belligerents. Because of its economic isolation during the war (due to the Allied blockade), the great reduction in its foreign trade, and its military expenditures in defense of its neutrality, upward pressures on prices were strong in the Netherlands. During the war nearly forty percent of the government's expenditures were debt-financed, and the government continued to run large deficits during the postwar years. The wholesale and retail price indexes increased nearly as much as those in the United Kingdom. Some price deflation was necessary for the return to prewar parity. Dutch prices fell continually from 1919 to 1923 and by the end of 1920 the guilder became undervalued on the basis of the retail indexes.

Monetary policy, at least as inferred from the level of the central bank's discount rate and frequency of changes in this rate, was not used extensively. With a large amount of liquidity resulting from wartime finance and the postwar deficits, the domestic impact of these changes in the discount rate did not appear to have been large. Instead, changes in the discount rate were in response to changes in the external value of the guilder. As the guilder appreciated the discount rate was lowered, from 4.5 percent to four percent in 1922, and from five percent to four percent in late 1924; as the guilder depreciated in late 1923, the discount rate was increased from four percent to five percent.

The absence of extensive upward price pressures, in face of the government deficits and the passive monetary policy, may be explained

in part by the relatively high level of unemployment. Registered unemployment among trade union members averaged about six percent in 1920, a boom year, and thereafter it was never below eight percent.

Achieving the prewar parity was somewhat easier for the Dutch than the British since the international position of the guilder had deteriorated to a lesser extent during the war; the Dutch had not been obliged to sell their foreign investments or to incur large foreign indebtedness. Their trading position, moreover, also seemed to have been less adversely affected. Nevertheless, their policy dilemma was similar to that of the British—should more expansive policies be pursued to reduce the level of unemployment, perhaps at the expense of lesser progress toward returning to the prewar gold parity? The Dutch, like the British, sacrificed a high level of employment in order to defend an overvalued exchange rate.

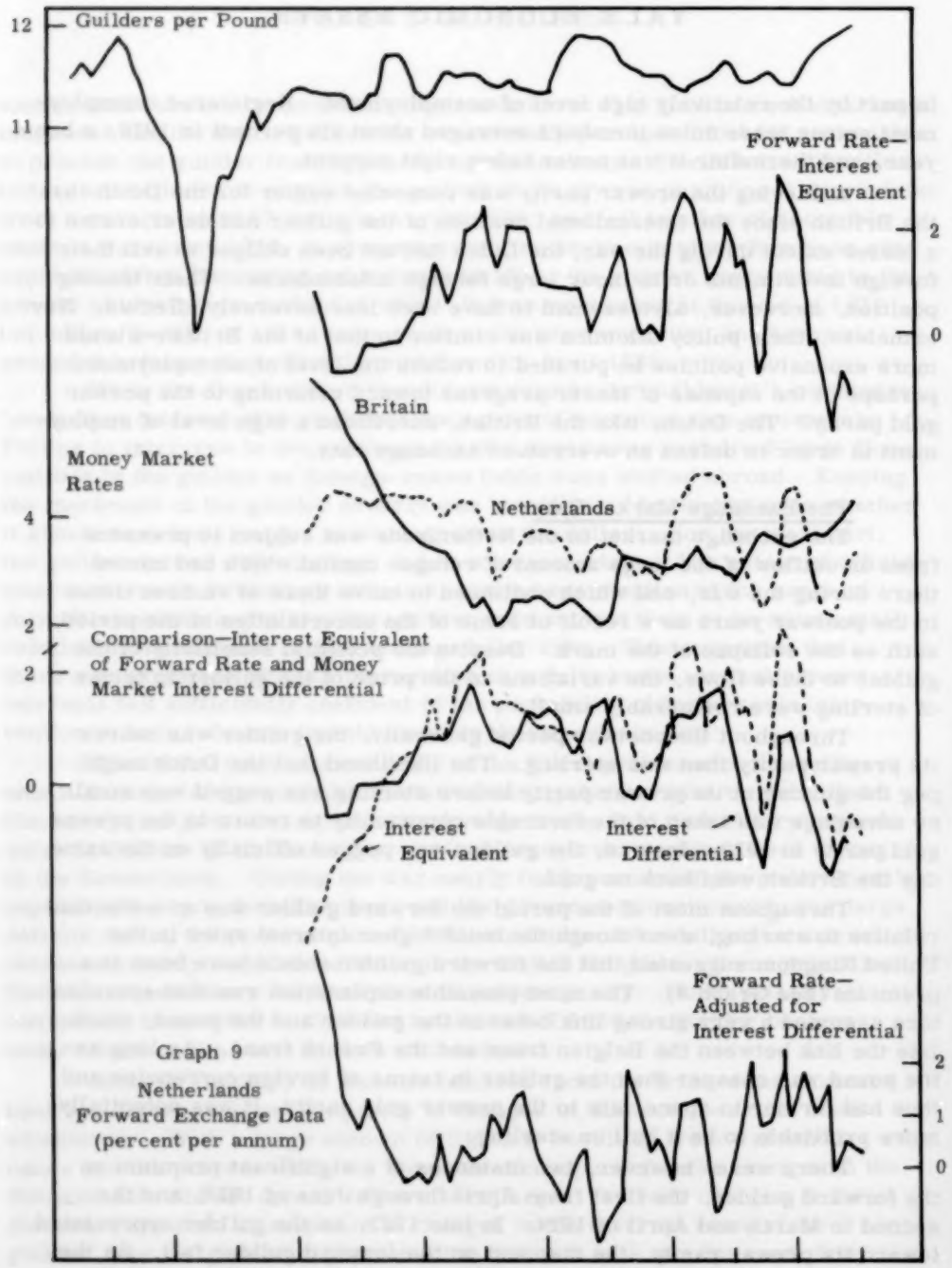
The Exchange Market Data

The exchange market in the Netherlands was subject to pressure from an outflow of the large amount of refugee capital which had moved there during the war, and which continued to move there at various times in the postwar years as a result of some of the uncertainties of the period, such as the collapse of the mark. Despite the potential sensitivity of the guilder to these flows, the variations in the price of the guilder in terms of sterling were remarkably small.

Throughout the postwar period generally, the guilder was nearer its prewar parity than was sterling. The likelihood that the Dutch might peg the guilder at its prewar parity before sterling was pegged was small; no advantage was taken of the favorable opportunity to return to the prewar gold parity in 1922. Instead, the guilder was pegged officially on the same day the British went back on gold.

Throughout most of the period the forward guilder was at a discount relative to sterling, even though the much higher interest rates in the United Kingdom suggested that the forward guilder should have been at a premium (see Graph 9). The most plausible explanation was that speculators assumed a very strong link between the guilder and the pound, much like the link between the Belgian franc and the French franc. As long as the pound was cheaper than the guilder in terms of foreign currencies and thus had further to appreciate to the prewar gold parity, it was potentially more profitable to be a bull on sterling.

There were, however, two instances of a significant premium on the forward guilder, the first from April through June of 1923, and the second in March and April of 1924. In late 1922, as the guilder appreciated toward its prewar parity, the discount on the forward guilder fell. As the guilder reached its prewar parity, the forward guilder moved to a small premium. As the guilder began to depreciate in April of 1923, a large forward premium developed. Forward speculators apparently were attempting to maintain the guilder at its prewar parity against the pressure of other market forces. The forward premium on the guilder in the spring



Graph 9

Netherlands
Forward Exchange Data
(percent per annum)

of 1924 occurred as the guilder was depreciating slightly relative to sterling. The premium at this time did not appear to be related to a sizeable movement in the rate or to other observable events.

Changes in the level of the forward discount on the guilder generally reinforced the explanation based on its link with sterling. In 1921, as the guilder depreciated against sterling, the discount on the forward guilder declined. This tendency was observable again in 1923 and 1924. By becoming less bearish on the guilder as it became weaker relative to sterling, forward speculators tended to limit the extent of fluctuations in the sterling-guilder rate.

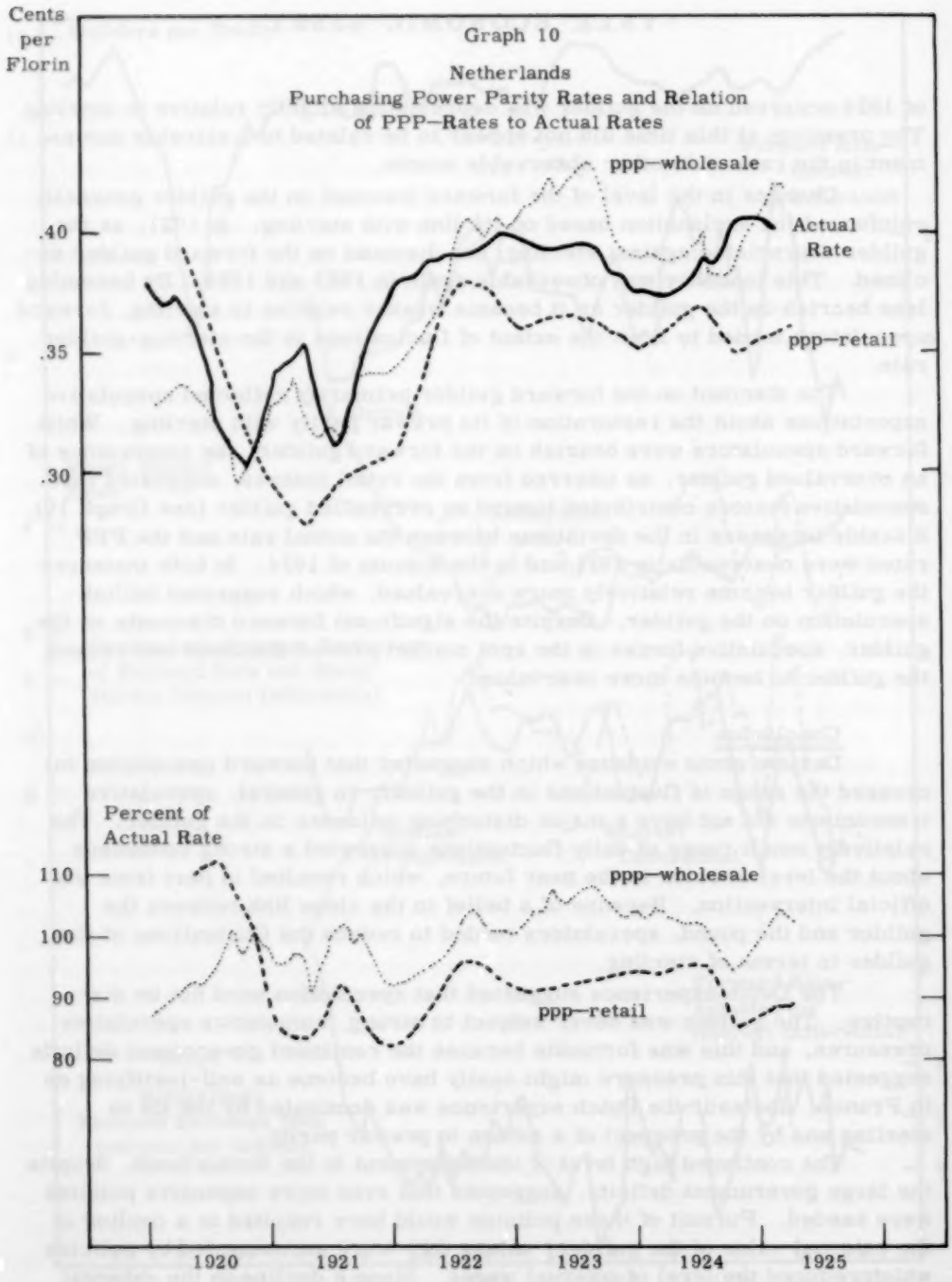
The discount on the forward guilder primarily reflected speculative expectations about the restoration of its prewar parity with sterling. While forward speculators were bearish on the forward guilder, the continuance of an overvalued guilder, as inferred from the retail indexes, suggested that speculative factors contributed toward an overvalued guilder (see Graph 10). Sizeable increases in the deviations between the actual rate and the PPP rates were observable in 1921 and in the autumn of 1924. In both instances the guilder became relatively more overvalued, which suggested bullish speculation on the guilder. Despite the significant forward discounts on the guilder, speculative forces in the spot market proved dominant and caused the guilder to become more overvalued.

Conclusion

Despite some evidence which suggested that forward speculation increased the range of fluctuations in the guilder, in general, speculative transactions did not have a major disturbing influence on the guilder. The relatively small range of daily fluctuations suggested a strong consensus about the level of rates in the near future, which resulted in part from the official intervention. Because of a belief in the close link between the guilder and the pound, speculators tended to reduce the fluctuations of the guilder in terms of sterling.

The Dutch experience suggested that speculation need not be disruptive. The guilder was never subject to strong, cumulative speculative pressures, and this was fortunate because the continued government deficits suggested that this pressure might easily have become as self-justifying as in France. Instead, the Dutch experience was dominated by the tie to sterling and by the prospect of a return to prewar parity.

The continued high level of unemployment in the Netherlands, despite the large government deficits, suggested that even more expansive policies were needed. Pursuit of these policies would have resulted in a decline in the external value of the guilder, unless they were accompanied by policies which reduced the level of internal wages. Since a decline in the external value of the guilder would probably have led to higher internal prices, the guilder might have been subject to increased speculative pressure. In this sense, the high unemployment level was the buffer which facilitated the return of the guilder to its prewar parity.


Cents per Florin
Graph 10
Netherlands
Purchasing Power Parity Rates and Relation of PPP—Rates to Actual Rates
ppp—wholesale
Actual Rate
ppp—retail
.40
.35
.30
Percent of Actual Rate
110
100
90
80
ppp—wholesale
ppp—retail
1920
1921
1922
1923
1924
1925

8. The Swiss Franc, 1919–1925

Switzerland was the only member of the Latin Currency Union to return to its prewar gold parity; it was also the only member of this group which remained neutral during the war. It experienced many of the same wartime dislocations as the Netherlands: large military expenditures, difficulty in obtaining needed imports, and a loss of traditional export markets. Moreover, it was forced to extend large export credits to both groups of belligerents.

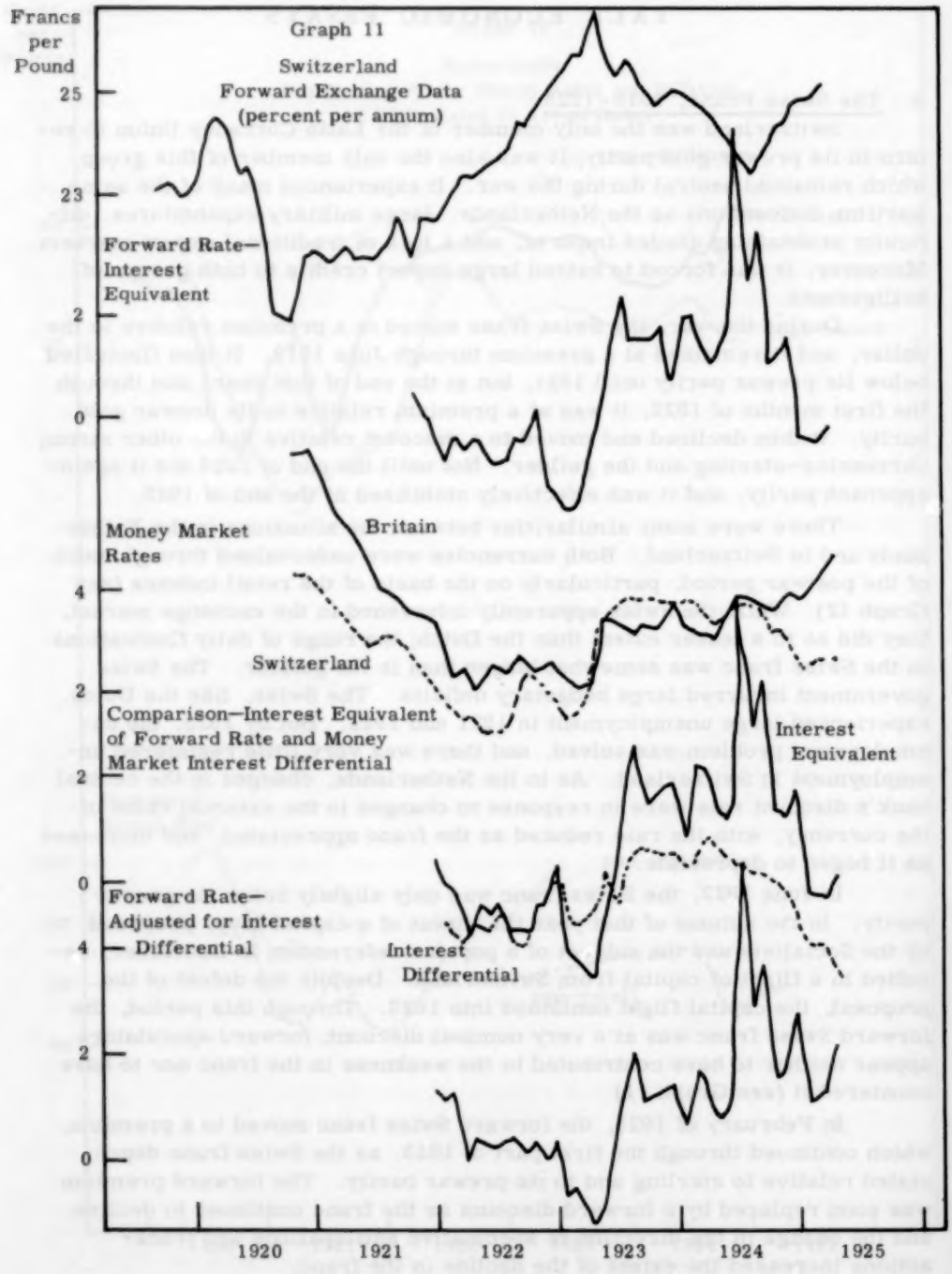
During the war, the Swiss franc moved to a premium relative to the dollar, and it remained at a premium through June 1919. It then fluctuated below its prewar parity until 1921, but at the end of that year, and through the first months of 1922, it was at a premium relative to its prewar gold parity. It then declined and moved to a discount relative to the other strong currencies—sterling and the guilder. Not until the end of 1924 did it again approach parity, and it was effectively stabilized at the end of 1925.

There were many similarities between the situations in the Netherlands and in Switzerland. Both currencies were undervalued through much of the postwar period, particularly on the basis of the retail indexes (see Graph 12). While the Swiss apparently intervened in the exchange market, they did so to a lesser extent than the Dutch; the range of daily fluctuations in the Swiss franc was somewhat larger than in the guilder. The Swiss government incurred large budgetary deficits. The Swiss, like the Dutch, experienced large unemployment in 1921 and 1922. But by 1923, the unemployment problem was solved, and there was very little registered unemployment in Switzerland. As in the Netherlands, changes in the central bank's discount rate were in response to changes in the external value of the currency, with the rate reduced as the franc appreciated, and increased as it began to depreciate.

In June 1922, the Swiss franc was only slightly below its prewar parity. In the autumn of that year the threat of a capital levy, proposed by the Socialists and the subject of a popular referendum in December, resulted in a flight of capital from Switzerland. Despite the defeat of the proposal, the capital flight continued into 1923. Through this period, the forward Swiss franc was at a very nominal discount; forward speculators appear neither to have contributed to the weakness in the franc nor to have countered it (see Graph 11).

In February of 1923, the forward Swiss franc moved to a premium, which continued through the first part of 1923, as the Swiss franc depreciated relative to sterling and to its prewar parity. The forward premium was soon replaced by a forward discount as the franc continued to decline, and the change in the direction of speculative anticipations and transactions increased the extent of the decline in the franc.

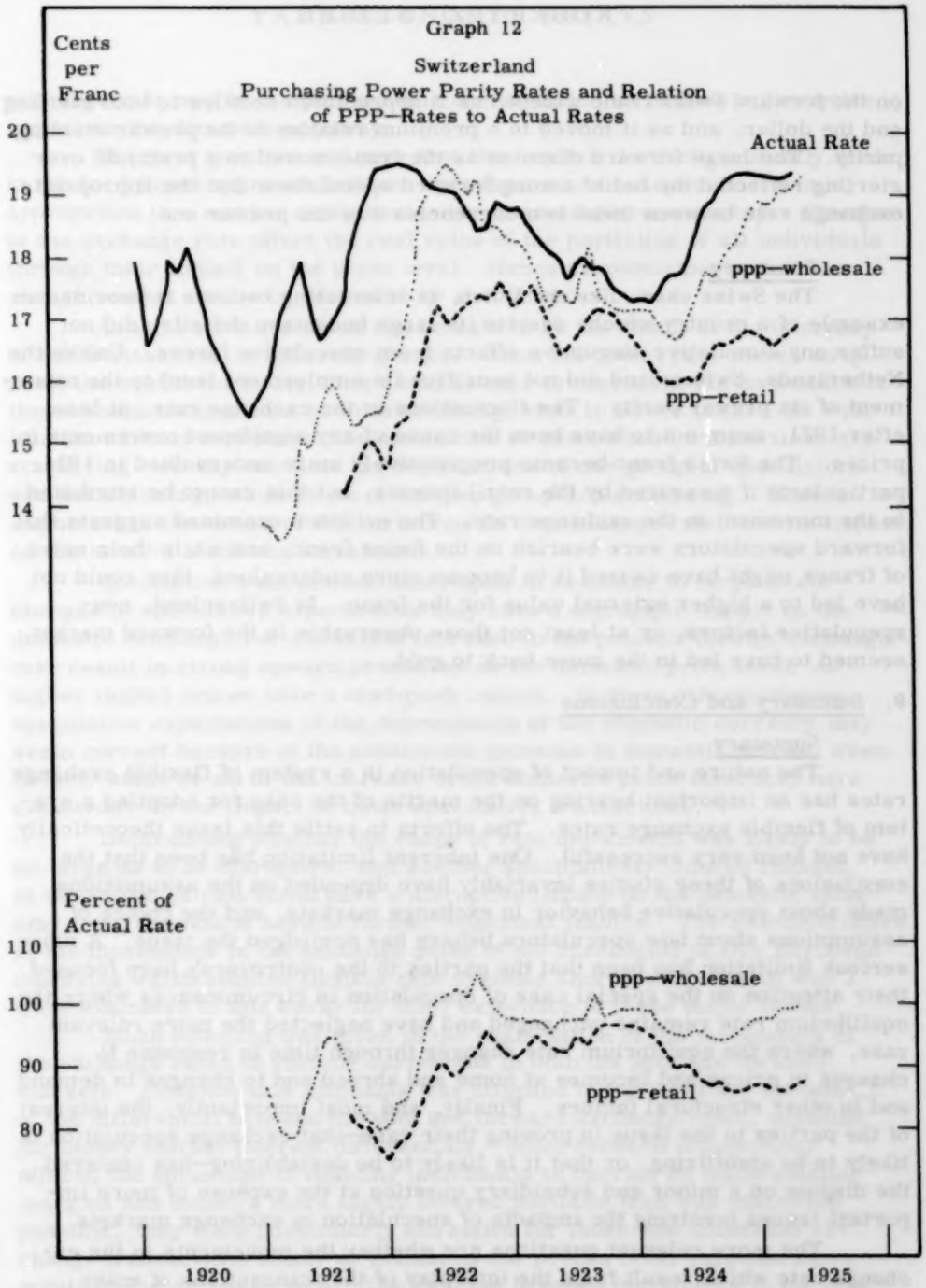
The discount on the forward franc remained slightly in excess of one percent through the second half of 1923 and until July 1924, while the spot franc weakened against the dollar and, to an even greater extent, against sterling. Then in August, September, and October, the discount


Francs per Pound
Graph 11
Switzerland
Forward Exchange Data
(percent per annum)
25
23
Forward Rate—Interest Equivalent
2
0
Money Market Rates
Britain
4
Switzerland
2
Comparison—Interest Equivalent of Forward Rate and Money Market Interest Differential
Interest Equivalent
2
0
Forward Rate—Adjusted for Interest Differential
4
Interest Differential
2
0
1920
1921
1922
1923
1924
1925

Graph 12

Switzerland

Purchasing Power Parity Rates and Relation of PPP—Rates to Actual Rates



on the forward Swiss franc widened as it appreciated relative to both sterling and the dollar, and as it moved to a premium relative to its prewar sterling parity. The large forward discount as the franc moved to a premium over sterling reflected the belief among forward speculators that the appropriate exchange rate between these two currencies was the prewar one.

Summary

The Swiss case, like the Dutch, is interesting because it provides an example of a country which, despite its large budgetary deficits, did not suffer any cumulative disruptive effects from speculative forces. Unlike the Netherlands, Switzerland did not sacrifice its employment level to the attainment of its prewar parity. The fluctuations in the exchange rate, at least after 1921, seem not to have been the cause of any significant movements in prices. The Swiss franc became progressively more undervalued in 1924, particularly if measured by the retail indexes, but this cannot be attributed to the movement in the exchange rate. The evidence examined suggests that forward speculators were bearish on the Swiss franc, and while their sales of francs might have caused it to become more undervalued, they could not have led to a higher external value for the franc. In Switzerland, non-speculative factors, or at least not those observable in the forward market, seemed to have led to the move back to gold.

9. Summary and Conclusions

Summary

The nature and impact of speculation in a system of flexible exchange rates has an important bearing on the merits of the case for adopting a system of flexible exchange rates. The efforts to settle this issue theoretically have not been very successful. One inherent limitation has been that the conclusions of these studies invariably have depended on the assumptions made about speculative behavior in exchange markets, and the choice of assumptions about how speculators behave has prejudged the issue. A more serious limitation has been that the parties to the controversy have focused their attention on the special case of speculation in circumstances where the equilibrium rate remains unchanged and have neglected the more relevant case, where the equilibrium rate changes through time in response to changes in prices and incomes at home and abroad and to changes in demand and in other structural factors. Finally, and most importantly, the interest of the parties to the issue in proving their case—that exchange speculation is likely to be stabilizing, or that it is likely to be destablizing—has centered the dispute on a minor and subsidiary question at the expense of more important issues involving the impacts of speculation in exchange markets.

The more relevant questions are whether the movements in the exchange rate which result from the interplay of the transactions of many speculators, who have different expectations about future rates, are likely to be sufficiently large so that they prove disruptive to international trade and investment, and whether the speculatively induced movement in the

exchange rate may be so pervasive in one direction as to have a disruptive impact on the domestic economy.

The first question assumes that the implicit distinction usually made in the traditional approach between speculators and nonspeculators is inappropriate in a market characterized by unavoidable uncertainty. Changes in the exchange rate affect the real value of the portfolios of all individuals through their impact on the price level. Hence all participants in the exchange market can be construed to be speculators.

Actual rate movements can be said to be the result of the interplay of the transactions of different groups. When the actual rate moves away from the equilibrium rate, this could be said to represent the dominance of those whose transactions are destabilizing; when it moves toward the equilibrium rate, it represents the dominance of those whose transactions are stabilizing. At any one time the transactions of one group of speculators are destabilizing, while those of another group are stabilizing. What is important is the extent of the variations in the exchange rate, and its impact on the level of trade and investment.

Speculative transactions undertaken in response to large-scale changes in speculative expectations may have a disruptive impact on the domestic economy, for a sizeable increase in the price of foreign exchange may result in strong upward pressures on the domestic price level, as higher import prices have a cost-push impact. In these circumstances, speculative expectations of the depreciation of the domestic currency may seem correct because of the subsequent increase in domestic prices, when, in fact, some or all of the increase in the domestic price level may have resulted from the impact of these speculative transactions.

Determining whether the range of rate movements was likely to be so large as to be disruptive, and whether speculatively induced changes in the exchange rate could have a disruptive impact on the domestic economy, was deemed in need of further empirical inquiry. The characteristics of the movements in the exchange rates of the currencies of five European countries which adopted flexible rate systems shortly after World War I were examined in this essay for their relevance to these major issues.

Much attention was given to the examination of the movements of the exchange rates of various currencies in both the spot and forward markets. Considerable emphasis was attached to the level and movement of the differential between the spot and forward exchange rates, adjusted for money market interest differentials. Since forward transactions offered the advantage of enabling individuals to operate on much smaller margins and obtain a more highly levered portfolio than was otherwise possible, they were particularly attractive for those who undertook exchange transactions solely or primarily for profits from expected movements in the exchange rate. The relation between movements in the spot rate and movements in the forward rate was examined to see whether it confirmed that the transactions of speculators in the forward market amplified the range of movement in the spot rate, consistent with the

overtracking hypothesis of speculative behavior.

The movements in the spot rate also were examined, and a time series of the ratio of the average range of daily fluctuation to the average foreign exchange value of the currency was developed. This ratio reflected the elasticity of the supply of short-term speculative funds; relatively low values for the ratio indicated strong speculative support for a currency. The movements in this rate were evaluated in relation to the movements in the differential between the spot and the forward rates.

The inferences about the speculative impact on the exchange rate based on the examination of forward market data were compared with inferences on the speculative impact based on the relation of the actual rate to a derived equilibrium rate. The Purchasing Power Parity technique was used to obtain a proxy for the equilibrium rates; it provided a check on whether the goods produced in countries with depreciating currencies were becoming less competitive with foreign goods. Depreciation of the domestic currency greater than that warranted by changes in the price levels at home and abroad suggested speculative pressure on the domestic currency.

Changes in economic variables likely to have a significant impact on the price level were examined in an attempt to determine the extent to which exchange rate movements followed the movements in the price levels, and the extent to which they may have caused these movements. Since government deficit financing was frequently cited as the cause of the rapid price increases in some countries, the price movements in various countries were considered in relation to the government's financial position as well as in relation to the exchange rate movements.

Conclusions

One of the main conclusions of the study was that the exchange rate in the flexible exchange market was not very resistant to shocks of various types, such as those resulting from political crises and temporary economic dislocations. The movement in the exchange rate could be sizeable and abrupt in response to factors which in retrospect were temporary and which frequently centered on foreign economies. Resistance to such shocks was relatively greater in the countries which had achieved a considerable measure of price stability. When shocks were severe and continuing, even the countries which had achieved relative price stability experienced a considerable depreciation of their currencies. The movement in the sterling-dollar exchange rate in the course of a day was as much as one or two percent at certain times. In late 1923, sterling fell about six percent within three months, or at annual rate of twenty-five percent, in reaction to a shock. The Dutch guilder and the Swiss franc declined significantly in reaction to shocks affecting their own countries.

Speculators in the forward market appeared to give little identifiable support to currencies which, because of shocks, were depreciating in the markets; the premium on forward sterling in the fall of 1923 seemed related to the underlying anticipation of sterling's eventual return to its prewar

parity. The support which the suspect currencies did receive was in the spot market, partly in response to the increasing differential between the price of domestic goods and the price of foreign goods which resulted from the movement in the exchange rate.

Speculators who used the forward market showed relatively little flexibility as a group, particularly the flexibility which would be required if they were expected to dampen exchange rate movements. Their acquisitions of weak currencies were almost always undertaken on the premise that these currencies would ultimately return either to their prewar gold parity, or to the prewar parity with another currency. There was very little evidence which suggested that forward speculators as a group increased their long positions in currencies which were temporarily weak, either in response to seasonal or other cyclical fluctuations or in response to shocks. It is significant that little evidence can be found which suggests that the group of speculators who strive for speculative profits dampened exchange rate fluctuations. It also is significant that, for an extremely long period, their expectations appeared based on some relationships established earlier, such as the prewar parities, rather than on the current factors which might have suggested that these objectives were beyond the reach of probable government policies.

The evidence that speculators in the forward market caused the actual rate movement to overtrack, as if they always followed a strategy of selling depreciating currencies and buying appreciating currencies, was limited. It was most apparent when there was strong speculative sentiment for a change in the level of the rate. In sterling, behavior consistent with the overtracking hypothesis was observed only toward the end of 1924, as spot sterling was appreciating toward its prewar parity. In the French franc and the Belgian franc, behavior consistent with this hypothesis was observed for an extended period when they were under considerable speculative pressure. In general, however, forward speculators did not appear to take their positions in accordance with this simple mechanical formula in the absence of strong speculative anticipations that the actual rate would move toward a different level.

In both strong and weak currencies, speculators caused the level of the actual rate to deviate from the equilibrium rate inferred by using the Purchasing Power Parity technique. On the basis of beliefs that sterling would return to its prewar gold parity, speculators eventually bid it toward this level and in doing so they caused it to become greatly overvalued. Their anticipations were encouraged by the official commitments to return to gold at the prewar parity. In the absence of this commitment, speculators would have had much less reason to be bullish on sterling.

By pushing sterling toward its prewar parity when this clearly was not justified by the cost-competitiveness of British goods, speculators forced the U. K. authorities to honor their commitment to return to the gold standard at the prewar parity when this was not the path of economic wisdom. The U. K. authorities were not willing to adopt the deflationary measures

necessary to secure this objective; instead, they hoped that it could be accomplished by an increase in the U.S. price level. When the initiative taken by speculators forced the sterling rate to its prewar parity, the U.K. authorities had to adopt extremely restrictive measures to maintain a greatly overvalued rate.

Speculators caused the French franc to become extremely undervalued after 1923. The value of the French franc was much below the level justified by the cost-competitiveness of French goods. Despite the extreme undervaluation, large-scale speculative sales of French francs continued on the presumption that the exchange value of the French franc would decline further. This decline was largely a result of speculation, and the adverse impact of speculation on government finances, rather than of government finances themselves. The evidence strongly suggested that the internal government policies adopted in late 1923 and early 1924, when the French franc already was substantially undervalued, would have been adequate to prevent further upward pressures on the French prices.

The Belgian franc also became extremely undervalued as a result of speculative pressures. The strong speculative belief that the Belgian franc and the French franc should exchange on a one-for-one basis made it impossible for the Belgians to pursue a policy of external stabilization successfully, even though they had been successful in achieving internal stability. The very rapid fall in the external value of the Belgian franc in the spring of 1926, after six months of internal price stability achieved on the basis of a pegged Belgian franc, indicated the difficulties that can beset an open economy when its currency comes under speculative attack.

A change from bullishness to bearishness in net forward speculative expectations generally had a major impact on the foreign exchange value of a currency. Such changes were infrequent and they did not occur instantaneously; yet, when they occurred, a very sizeable movement in the exchange rate resulted. In the case of the French franc, a change of this sort in late 1923 was one of the principal factors which led to the decline in the external value of the French franc of about fifty percent in twenty weeks. Speculative bullishness about the franc may have been misplaced earlier and an eventual change in expectations not to be unexpected. Nevertheless, the change, when it finally came, had an extremely powerful impact on the value of the franc.

Abrupt movements in the exchange rate occurred as speculation "became all one way"—as the expectations of many speculators changed at the same time. While such changes tended to induce responses within the domestic economy which tended to limit the movement, these were not strong. The first check or restraint to a continued decline in the exchange rate came from the changing price differentials between similar goods produced or available domestically and abroad, which resulted from the decline in the exchange rate. The second check or restraint came from the contractive impact of higher domestic prices on the domestic interest rate structure, as the need for money balances increased. Because of the lags in changes in demand in response to changing price differentials, the first restraint operated with considerable delay. To the extent that the public already had

excess money balances, or could easily obtain them, the potency of the second restraint was weakened. A large, disruptive movement in the exchange rate could occur in the short-run before either of these automatic limiting forces became effective.

When speculators bid up the price of foreign exchange, the cost of imports increased; higher import prices exerted a cost-push impact on the domestic price level. The inflationary impact of bearish speculative transactions operating primarily from the cost-push side became the major source of upward pressure on prices and costs in France and Belgium after the end of 1922, and probably as early as the middle of that year. While the government deficits may have contributed toward the increase in prices through demand-pull pressures, various estimates based on the available data suggested that the upward pressures on prices from this source was much less than that from the demands associated with generating an export surplus necessary to finance the speculative capital outflow. Fiscal deficits in France in 1924 and thereafter were small and much smaller than the estimates of the capital outflow.

France and Belgium were particularly susceptible to self-aggravating speculative disturbances because of the regressive nature of their tax systems. In periods of rising prices, an increase in tax rates was necessary to reduce the almost automatic increase in the government's deficit. Yet, it was exceedingly difficult to raise tax rates sufficiently in the absence of a consensus on which taxes should be raised. Nevertheless, achievements in reducing government deficits must be considered quite impressive; after 1923 Belgium had virtually ceased to run deficits, and France incurred only small ones.

Moreover, the upward pressure on interest rates which might have resulted from the increased demand for money balances in a period of rising prices was weakened by the encashment of large holdings of short-term government securities, so that any tendency toward a reversal of speculative expectations which might have resulted from upward movements in interest rates was limited. Further, the conversions of these securities greatly complicated the management of the Treasury debt and stimulated expectations of even more inflation.

Since the evidence suggested, however, that most speculators did not operate on fine margins, it seems unlikely that the increases in the interest rate which might have resulted even if the money supply had been fixed would have been sufficient to accomplish a reversal of speculative expectations before a rapid increase in prices had occurred, and before subsequent wage boosts had become ratified. Speculatively induced changes in the exchange rate occurred suddenly, and there was an extended lag before there was an equivalent change in domestic prices and a sizeable increase of interest rates. The problem of financial management, as long as bearish speculation led to higher domestic prices, was to reverse speculative anticipations. Something much more potent than an upward change in the interest rate structure was needed to prevent speculation from feeding on itself.

Moreover, because of the lag before these checks to a continued decline in the exchange rate became effective, even countries which had pursued much more restrictive financial policies could have experienced a large decline in the exchange value of their currencies if speculative expectations suddenly changed.

Both the Netherlands and Switzerland avoided the circle of depreciating currencies and rising prices and eventually returned to their prewar gold parities. Their experience is sometimes cited to prove that exchange speculation need not be disruptive. One of the critical factors in their favor was the extensive intervention of their authorities in the exchange markets to keep rate movements from becoming sizeable or disorderly. In many other respects, however, the similarities of economic developments between France and Belgium, and the Netherlands and Switzerland, were more striking than the differences. In the Netherlands and Switzerland, the governments incurred large fiscal deficits, and a large part of new debt sold by the government was short term. Their tax systems also were regressive. Both countries, therefore, easily could have become subject to a speculative attack which proved self-justifying.

Speculative expectations about the Dutch guilder and the Swiss franc were strongly influenced by the belief that they should exchange at their prewar parity with sterling. As long as this expectation, which was comparable to the speculative belief in the linkage between the French and Belgian francs, proved dominant, there was little likelihood of a speculative attack. If it had disappeared, and if intervention had not been adequate to cope with speculative pressures, the history of the movement of the guilder and the Swiss franc might have been more nearly like that of the French and Belgian francs.

The European experiences with flexible rate systems were dominated by many factors which might be construed unique to the period, the most dominant being the anticipations of the return to prewar parity with gold or with another currency. These anticipations slighted the real difficulties in re-establishing the prewar world; the international competitive position of most European countries had deteriorated during the war. Together with the loss of foreign investments and the increase in foreign indebtedness, this made the return to the prewar parities all the more difficult. The need to make up for the economic dislocations resulting from the war—the destruction, depleted inventories, deteriorated plant and equipment—affected the level of post-war expenditures and contributed to a year of very rapid price increases in all countries beginning in the spring of 1920. The continuing struggles over reparations among the belligerents and over war debts among the allies eventually affected expectations of the future rates. The lack of a broad political consensus in many countries, the widespread political instability, and the occasional threats of a capital levy increased the sensitivity of the owners of capital to changes in domestic policies. All these factors affected the functioning of the exchange markets, and most of them complicated the return to parity.

Yet the impact of most of these factors was appreciated by speculators only with great delay.

The conclusions about the characteristics of exchange speculation might have been somewhat different if they had been derived from situations of greater political stability. Political uncertainty and instability exist, however, and the evaluation of the desirability of an exchange rate system must recognize how it adjusts to them; it must also recognize whether it may contribute to and accentuate them.

The expectation of a return to the prewar parities probably reduced the range of rate fluctuation. Nevertheless, reliance on market forces and the interplay of speculative transactions for the determination of the actual rate tended to result in a relatively large range of rate fluctuations, both on a daily basis and over more extended periods. After the expectation of a return to prewar parities had disappeared, there was little evidence to suggest that speculators could be relied on to adjust the actual rate to changes in the equilibrium rate as they occurred. Instead, the more the currencies depreciated, the more bearish speculators became about their future. Speculators showed little of the flexibility which would be required to keep rate movements small, orderly, and in line with changes in the equilibrium rate.

The case for the flexible rate system appears attractive until the question of whether rate movements are likely to remain small and orderly, both in the short run and in the long run, is examined. The evidence examined suggests that this cannot be left to market forces—that if it is, the rate movements are likely to be large, and even disorderly, in the short run. In the long run, reliance on market forces for determination of the exchange rate places the system on the brink of a potential disaster because of the rapidity with which changes in the exchange rate, in response to a change in speculative expectations, can exert a self-justifying inflationary impact on the domestic economy before automatic, limiting forces can become effective and before new government policies can be introduced.

Summary: This essay examines the changes which have been experienced by the United States over the first half of the twentieth century in two major economic variables, and the implications of these changes with respect to the portfolio choices of the private sector of the economy. These changes are the decline in the producer capital coefficient and the fall in the ratio of the national debt to the net national product. For purposes of analysis, a model is constructed which assumes the demand for stocks of wealth assets to be a function of income. The model is used to examine the paths of adjustment which arise when there is a discrepancy between the quantity of wealth assets demanded by the public and the quantity supplied by producers and government. For the short run, the emphasis is on changes in the valuation of wealth assets, while for the long run, the changes in the wealth portfolio of the public and the capital intensity of production are emphasized. Implications of the analysis for monetary policy and for government debt policy are considered in the final section.

This essay is based on a dissertation which was submitted to Yale University in 1957 in partial fulfillment of the requirements for the degree of Doctor of Philosophy, which was awarded in 1958. The author is currently an economist in the Western Hemisphere Department of the International Monetary Fund.

Dissertation Committee: Professors William J. Fellner (Chairman), Challis A. Hall, and Charles E. Lindblom.

*

The author is greatly indebted to Professor William Fellner for his valuable assistance, advice, and criticism during the preparation of this paper, and to Professors Tjalling Koopmans, Jacob Marschak, Richard Ruggles, James Tobin, and Henry Wallich for their help on special points. The author wishes to thank Dr. Raymond Goldsmith for his advice on specific data in the latter's Study of Saving, and Mr. Kazuo Sato, who helped in the preparation of a mathematical appendix to the original dissertation. A special debt is owed Professor John Perry Miller for his assistance throughout the preparation of this study. The author accepts full responsibility for the contents of this paper and any errors that may appear herein.

R. S. T.

LONG-RUN ADJUSTMENTS TO CHANGES IN THE CAPITAL-OUTPUT RATIO AND THE NATIONAL DEBT

Richard S. Thorn

A. Introduction

This paper focuses attention on two economic problems which heretofore have received scant attention. They are the economic implications of the recent fall in the producer capital coefficient and the decline in the ratio of the national debt to the national product of the United States. The common characteristic of these disturbances is that they both require changes in the asset portfolio of the public.

The method followed in this paper is to introduce an identity relating the community's desire to hold a given amount of wealth per unit of income with producers' desire to employ a given amount of capital per unit of output. This identity provides the analytical framework within which the process of mutual adjustment caused by changes in the capital-output ratio and the wealth-income ratio over the first half of the 20th century is studied. The conclusion of the empirical study is that the capital coefficient of producers and the wealth-income ratio of non-farm households remained relatively stable until 1930. In the postwar period the capital-output ratio declined sharply but the wealth-income ratio of the private sector declined to a lesser degree as a result of the simultaneous increase in government debt.

On the assumption that the amount of wealth individuals desire to hold is a function of their income, a model is introduced to analyze the paths of adjustment that may occur when the amount of wealth claims desired by individuals per unit of income differs from the amount supplied by producers and the government. In the short run, the model emphasizes the importance of changes in the valuation of wealth assets, particularly intangible wealth claims such as securities, in restoring equilibrium. In the long run the emphasis is placed on changes in the wealth portfolio of the public and changes in the capital intensity of production.

In the face of wage rigidity and possible "capital shortages," the possibility exists that orthodox monetary policy alone will be unable to avert simultaneous inflation and unemployment. The final section of this paper raises the question as to whether a consideration of the assets effects of changes in government debt might lead to a more effective monetary policy, and whether or not the optimum rate of growth of government debt is zero.

There are two alternative ways of looking at economic processes. One is to trace the flows from one economic unit to another, and the other is to observe the change in the stocks of assets of the various economic units. These two views are analogous to the income and balance sheet statements of a private business firm. Each viewpoint has its own special advantages. In the history of economics these two aspects have been introduced into analysis in varying proportions depending on the problem at hand. In an attempt to achieve a closer synthesis of the real process of capital accumulation with the essentially monetary and financial process of the community's acquisition of wealth, the present paper continues the recent emphasis on the role of stocks of assets in economic processes.

B. The Basic Identities

Wealth (U) is defined as the sum of claims to producer capital goods (K), consumer capital goods (D), net corporate and non-corporate business equities (E), all other net financial assets (C), and net claims against foreign countries (F). When aggregated for the economy, all financial and equity claims total zero since for every asset there is an offsetting liability. The equity of business is regarded as an asset of its owner claimants and a liability of the business itself. Since all financial claims cancel one another, aggregate wealth (U) equals aggregate capital (Z).

$$(1)\quad U = K + D + E + C + F = K + D + F = Z$$

Consumer capital is arbitrarily defined as consisting of all consumer goods with a life of approximately three years or more, including residential housing. Inventories of consumer durables and non-durables in the hands of producers and sellers are considered producer capital.

Income (Y) is defined as the total output of goods and services net of current depreciation, i.e. net national product (NNP). The value of the services of consumer durables, with the exception of residential housing, is excluded from income.[1]

The aggregate wealth to income ratio (U/Y) is then defined as the sum of the producer capital to output ratio, the durable goods to output ratio, and the net foreign goods to output ratio, and is equal to the aggregate capital to output ratio (Z/Y).

It is useful to examine the same identity at a lower level of aggregation. Not only is it possible to distinguish the various economic groups that may exhibit significantly different behavioral patterns, but it is also possible to relate the sectoral relationships to the national aggregate. To accomplish this the following terms are introduced:

C_i^j = Net claims excluding equity claims of the i-th group against the j-th group; a negative value reverses the relationship.

E_i^j = Net equity claims of the i-th group against the j-th group; a negative value reverses the relationship.

D_i = Consumer durable goods of the i-th group (provided it is a consumer group).

K_i = Capital of the i-th group (provided it is a producer group).

T_i = Tax and non-tax receipts of the i-th group (provided it is a government group).

---

1. S. Kuznets estimates that the use value of consumer durables since 1870 has been between 3/4 and 1-3/4 per cent of gross national product. Cf. S. Kuznets, "Long-run Changes in the National Income of the United States of America since 1870," Income and Wealth, Series II (Cambridge: Bowes and Bowes, 1952), p. 165.

Y_i = Income after taxes of the i-th group (provided it is not a government group).

U_i = Wealth or net worth of i-th group ($U_i = K_i + E_i^j + D_i + C_i^j$)

X_k = Capital of the k-th industry.

V_k = Value added to national product net of depreciation by the k-th industry.

F = Net claims against foreigners.

A capital letter without any subscripts or superscripts indicates a net aggregate for the entire economy, unless otherwise defined. The net national product of the economy (Y) is equal to the sum of the income after taxes of the behavioral groups plus the sum of the tax and non-tax receipts of the government groups.

A relatively simple identity (2) can be constructed showing the interrelationship between the wealth-income ratios of the various behavioral groups and the capital-output ratios of the various industries. Identity (2a) distinguishes the behavioral groups and identity (2b) the industries.

$$(2)\quad \frac{U}{Y} = \frac{Z}{Y}$$

$$(2a)\quad \frac{U}{Y} = a + b + c + d;$$

$$\text{where}\quad a = \frac{K_1 + C_1^2 + C_1^3 + C_1^4 + C_1^f + E_1^2 + E_1^3}{Y_1} \cdot \frac{Y_1}{Y},$$

$$b = \frac{D_2 + C_2^1 + C_2^3 + C_2^4 + C_2^f + E_2^1}{Y_2} \cdot \frac{Y_2}{Y},$$

$$c = \frac{C_3^1 + C_3^2 + C_3^4 + C_3^f + E_3^1}{Y_3} \cdot \frac{Y_3}{Y},$$

$$d = \frac{K_4 + C_4^1 + C_4^2 + C_4^3 + C_4^f}{T_4} \cdot \frac{T_4}{Y}$$

$$2b)\quad \frac{Z}{Y} = \frac{X + D + F}{Y} = \frac{X_1}{V_1} \cdot \frac{V_1}{Y} + \cdots + \frac{X_k}{V_k} \cdot \frac{V_k}{Y} + \frac{D}{Y} + \frac{E}{Y}$$

Identity (2) and its components relate the amount of wealth held per unit of income to the capital stock existing per unit of output. The terms of (2a) are chosen on the basis of the homogeneity of economic behavior and are categorized by functional institutional criteria; the terms of (2b), on the other hand, are chosen largely on the basis of similarity of production function and homogeneity of capital-output ratios. The identity (2) and

its components (2a and 2b), considered as a whole, demonstrate the truism that the wealth-income ratio and the structure of claims are related to the capital-output ratio; any change in the elements of one requires mutual adjustment among other elements. It is possible, however, that adjustments may occur among the terms of only one member of the identity.

The various terms of (2a) characterize the economy as follows: term (a) refers to the productive group which holds the productive capital of the economy, and against which the other sectors hold debt and equity claims; term (b) refers to the consumer group which holds consumer durables, debt and equity claims against other groups, and against which debt claims are held; term (c) refers to the financial intermediary group whose assets consist entirely of positive and negative claims against other groups; and term (d) refers to the government group, which is distinguished from the other groups in that its receipts flow not from its holding of capital, but from its ability to tax. Each of these terms consists of two factors: the first is the intra-group wealth-income (or net worth-profit) ratio, and the second ratio represents an income distribution, which may be regarded as a weighting factor, based on the share of that particular group in total national income.

The treatment of the second component (2b) of the basic identity is analogous to that of (2a). The terms here represent industries or, if disaggregated further, firms. The two factors are an intra-industry, or firm, capital coefficient and a weight that represents the composition of output. A detailed development of this aspect of the identity will be omitted since, for purposes of this study, the aggregate of this member will be of more interest than its components.

The basic identity provides an accounting framework which permits disaggregation of the economy, or any particular section of it, to any degree that may be desired. Our empirical analysis will be conducted within this framework.

C. Historical Movements of the Aggregate[1] Capital-Output and Wealth-Income Ratios

Assuming that the observed values tended to equal the desired values in the long run, it is possible, by using identity (2), to observe the nature of the mutual adjustment that took place between the capital coefficient and the

1. Owing to limitations of space it is possible to present only summary results of the statistical investigation which was actually conducted. Detailed tables which give sectoral breakdowns by type of assets and other detailed evidence supporting the factual statements of this section may be found in R. Thorn, Long-Run Adjustments to Changes in the Capital-Output Ratio and the National Debt, unpublished doctoral dissertation (Yale University, 1958), Statistical Appendices, pp. 97–139. This will be referred to henceforth as R. Thorn. It should be noted, however, that the valuation of land in constant prices differs from that in the present study. Copies of the Statistical Appendices may be obtained upon request from the author.

wealth coefficient over the first half of this century. In Table 1, the aggregate capital-output ratio is presented in the form of decade-average annual ratios in 1929 prices. Table 2 presents the same data in terms of the current replacement cost of wealth. The capital and wealth concept is that of tangible assets (net of depreciation) plus net foreign assets, including government capital, with the exception of military assets; the income concept is that of net national product, using original cost of depreciation. [1]

The relationship between the aggregate ratios and their components, which are presented in Tables 1 and 2, may be derived from the following identities:

(3) National Wealth = Reproducible Tangible Wealth (RTW)
+ Unimproved Land + Net Foreign Assets
= National Capital

(4) RTW = Non-Residential Structures + Producer Durables
+ Residential Structures + Consumer Durables
+ Inventories (agricultural and non-agricultural)
+ Monetary Gold and Silver

Chart 1, which illustrates the data presented in Tables 1 and 2, clearly indicates that, with the exception of the post-World War II period, both the constant and current price series exhibited the same general movements. We may therefore confine our remarks to the constant price series without a separate discussion of the current price series. The post-war divergence of the two series will be discussed separately.

The decline in the aggregate capital-output ratio before 1929 can be accounted for almost exclusively by the fall in the land-output ratio, which in turn is largely attributable to the diminishing importance of agriculture in the economy. [2] During World War II, in addition to the downward trend

1. This introduces an inconsistency in the current price series which is based on replacement cost depreciation. The discrepancy is minor except for the post-World War II period, where this leads to slightly lower values, e.g., the aggregate wealth coefficient in current prices for 1949 would be increased from 3.79 to 3.98. This inconsistency was allowed to remain in order to use the detailed Department of Commerce data, which are calculated on an original cost basis.

2. Land is valued as a proportion of the value of the relevant structures. If the value of land is held constant at its 1929 price, as the author and some writers have done in the past, the aggregate wealth-income ratio would be: 1900-09, 6.1; 1910-19, 6.4; 1920-29, 5.2; 1930-39, 5.4; and 1940-49, 3.3. This latter method, however, clearly overvalues land for the first two decades. Cf. R. Goldsmith, A Study of Saving in the United States (Princeton: Princeton University Press, Vols. 1 and 2, 1955, Vol. 3, 1956 — henceforth referred to simply as R. Goldsmith), Vol. 3, p. 17.

TABLE 1

Average Annual Ratios of National Wealth and Selected Components to Net National Product in 1929 Prices, 1900 – 1952

Period	Ratios to Net National Product						
	National Wealth	Land	Reproducible Tangible Wealth	Non-Residential Structures and Producer Durables	Residential Structures and Consumer Durables	Inventories	Monetary Metals
1900–1909	5.12	1.69	3.43	1.60	1.36	.45	.08
1910–1919	4.88	1.44	3.44	1.66	1.33	.41	.04
1920–1929	4.84	1.31	3.51	1.60	1.39	.46	.06
1930–1939	5.21	1.36	3.85	1.76	1.59	.42	.08
1940–1949	3.10	.70	2.44	1.04	.98	.32	.10
1952[1]	3.14[2]	.64[2]	2.50	1.05	1.02	.33	.08
1900–1929	4.94	1.48	3.46	1.62	1.36	.44	.06
1900–1939	5.00	1.45	3.55	1.65	1.42	.43	.07
1900–1949	4.63	1.30	3.33	1.52	1.33	.41	.07
1947–1949	3.15	.63	2.52	1.08	1.02	.33	.10

1. Rough estimates.

2. Estimating land at its 1949 value.

Source: R. Thorn, Statistical Appendix 1. Data above differ from that in R. Thorn in that the value of land was estimated as proportional to the value of the relevant structures rather than held constant at its 1929 value.

TABLE 2

Average Annual Ratios of National Wealth and Selected Components
to Net National Product in Current Prices
1900 – 1952

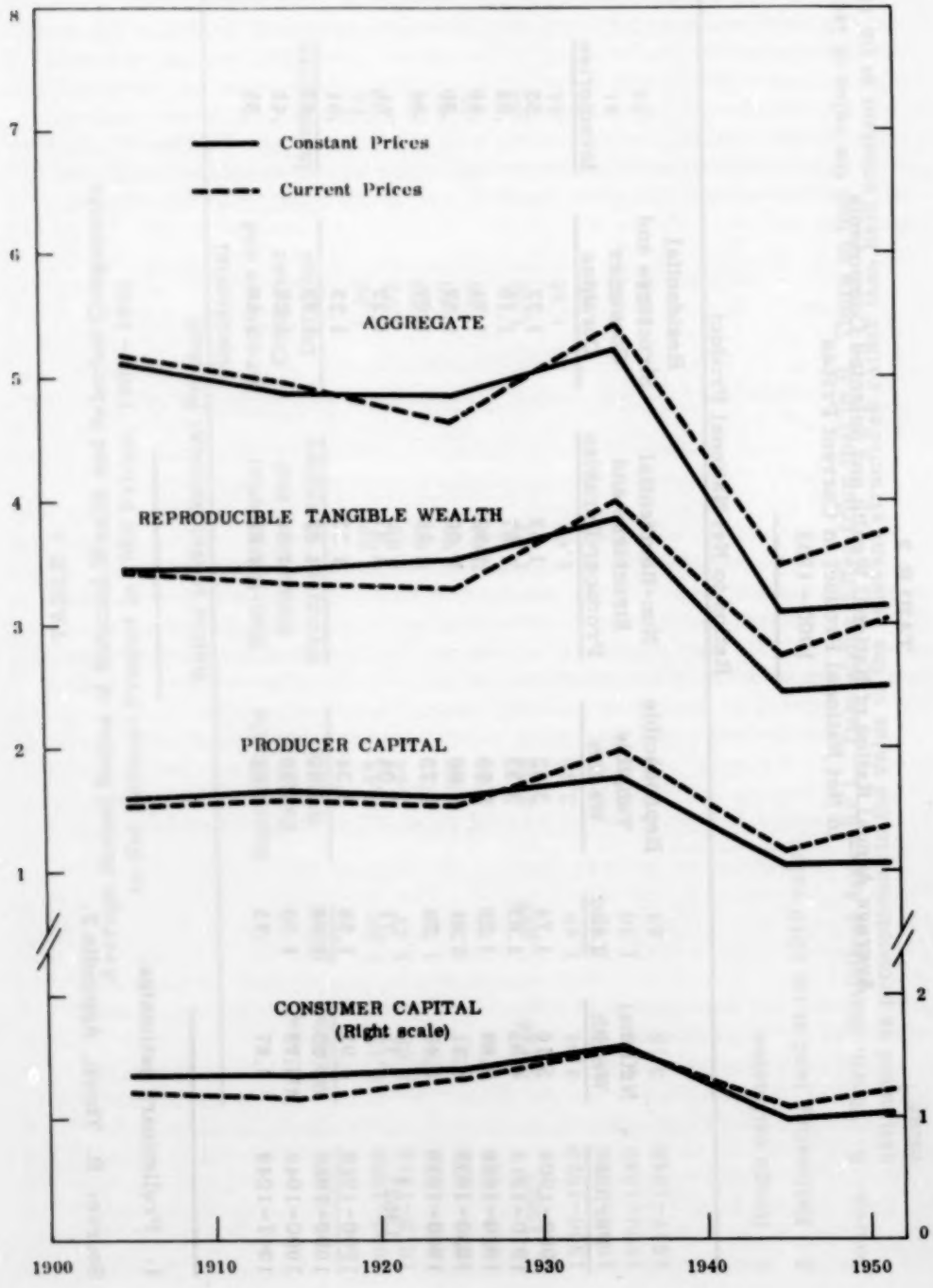
Period	Ratios to Net National Product						
	National Wealth	Land	Reproducible Tangible Wealth	Non-Residential Structures and Producer Durables	Residential Structures and Consumer Durables	Inventories	Monetary Metals
1900–1909	5.16	1.74	3.42	1.52	1.22	.53	.09
1910–1919	4.95	1.62	3.33	1.57	1.16	.52	.07
1920–1929	4.68	1.39	3.29	1.52	1.31	.49	.06
1930–1939	5.41	1.41	4.00	1.90	1.57	.39	.15
1940–1949	3.47	.74	2.73	1.18	1.07	.34	.16
1952[1]	3.75	.71	3.04	1.37	1.22	.35	.08
1900–1929	4.93	1.59	3.34	1.53	1.23	.51	.07
1900–1939	5.05	1.54	3.51	1.62	1.32	.47	.09
1900–1949	4.73	1.38	3.35	1.54	1.27	.45	.11
1947–1949	3.67	.75	2.92	1.27	1.18	.35	.12

1. Preliminary estimates.

Source: R. Thorn, Appendix 2.

Chart 1

THE AGGREGATE CAPITAL COEFFICIENT AND SELECTED COMPONENTS IN CONSTANT AND CURRENT PRICES



of the land ratio, a sharp fall occurred in the RTW ratio, caused by a decline in both its producer and consumer components.[1] In 1952 both RTW components were still significantly below their prewar levels in terms of constant prices, the low level of the consumer component resulting from the low level of the housing component. In terms of current prices, however, the RTW ratio, and all its components, had achieved its prewar level. This agreement would be even closer if NNP had been calculated on the basis of replacement cost depreciation.[2]

According to preliminary estimates for 1955 by R. Goldsmith, the producer component of the aggregate capital/NNP ratio in constant prices was still significantly below its prewar values.[3] From the latest data available (1957), the postwar trend in the RTW ratio cannot be determined with certainty. The RTW ratio rose steadily between 1945 and 1949, but the figure in 1952 was still approximately that of 1949. It is not possible to determine whether this was caused by the leveling off of the ratio, or by the rapid rate of income expansion which was taking place because of the Korean War. These data should soon be available to determine whether the upward trend in the RTW ratio, and particularly in its producer component, has continued.[4] In summary, it can be said that the downward trend in the aggregate capital-output ratio in constant prices, which occurred in the first quarter of the century, was caused by the fall in the land-output ratio. This tendency was increased in the second quarter of the century by an additional decline in the producer capital and residential housing components of RTW. The exact dating of the beginning of this latter downward trend in the aggregate ratio is somewhat open to question. If it did begin in the twenties in manufacturing and mining, as the National Bureau Studies seem to indicate, then it undoubtedly was accelerated during World War II. The National Bureau

---

1. There are strong indications based on recent studies by the National Bureau of Economic Research that the capital-output ratio in manufacturing and mining started its downward trend during the twenties, even though this is not shown in our decade averages and is only suggested slightly in the annual data. See D. Creamer, Capital and Output Trends in Manufacturing Industries, 1880-1948, and I. Borenstein, Capital and Output Trends in Mining Industries, 1870-1948 (New York: National Bureau of Economic Research, Occasional Papers Nos. 41 and 45, respectively, 1954).

2. However, the residential housing sub-component was still significantly below its prewar level in constant prices in 1949. The ratio of residential structures to disposable personal income in 1929 prices was: 1900, 1.2; 1929, 1.1; 1949, 0.8.

3. Cf. U.S. Bureau of Census, Statistical Abstract of the United States, 1957 (Washington, 1957), pp. 319-320.

4. See page 256.

data are consistent with R. Goldsmith's data if we are willing to assume that the capital coefficients for the United States during World War I were lower than those of the corresponding boom periods in the twenties, so that the second decade averages have a downward bias because of the abnormal conditions of the war. In spite of the differences between the two estimates it is clear that the post-World War II ratio was below that of any prewar period.

In current prices the same movements are observed, except that, by 1952, the RTW ratio and its components, with the exception of residential housing, seemed to have attained their prewar levels.

Since the writing of this paper in the summer of 1957, additional estimates of national wealth were prepared by Dr. Goldsmith for the years 1945, 1950, 1955 and 1958. These estimates are not wholly comparable with those given in Tables 1 and 2 of this study because of a change in the base period from 1929 to 1947-49 and other minor changes in the sources of data and methods of estimation used. Subject to this qualification, estimates comparable to Tables 1 and 2 have been prepared and are given in Table 3. These recent data support the conclusions drawn on the basis of preliminary estimates for 1952 and 1955. The various ratios of components of national wealth to net national product have continued to approach the long-run historical values shown in Tables 1 and 2 with the exception of the land and inventory ratios. In most cases both constant and current price ratios had attained values close to their long-run average values by 1958.

D. The Historical Intersectoral and Intrasectoral Wealth-Income Adjustment

Corresponding to changes in the aggregate capital/wealth-income ratios, a process of mutual adjustment was occurring in both the intrasectoral wealth-income ratios and in the weighting of the various sectors. These changes will be analyzed in the framework of identity (2).

Four final behavioral sectors are distinguished:

1) Non-Farm Households, which for convenience include also non-financial, non-profit institutions and personal trust departments.[1]

---

1. There is a serious inadequacy in our sectoring of the non-farm households. It is well known that the major portion of corporate equities is held by upper income groups. In the Survey of Consumer Finances for 1950 it was estimated that 75 per cent of individually owned corporate securities were held by individuals with incomes of $7500 or more, who thus received approximately 20 per cent of personal income. See R. Goldsmith, Vol. III, p. 126. A minimum adequate sectoring would distinguish the upper income groups who achieve a substantial part of their saving through corporate equities from the other groups who save largely through other forms of wealth. Unfortunately the data needed to perform this sectoring adequately are not available at present.

TABLE 3

Annual Ratios of National Wealth and Selected Components to Net National Product 1945 - 1958

Period	National Wealth	Land	Reproducible Tangible Wealth	Non-Residential Structures and Producer Durables	Residential Structures and Consumer Durables	Inventories	Monetary Metals
				Current Prices			
1945	2.84	.57	2.27	.95	.94	.26	.12
1950	3.98	.72	3.22	1.38	1.37	.36	.10
1955	3.79	.65	3.09	1.39	1.32	.31	.07
1958	4.14	.72	3.37	1.58	1.40	.32	.06
				1947-1949 Constant Prices			
1945	-	-	2.51	1.10	1.03	.30	.09
1950	-	-	3.10	1.32	1.33	.35	.10
1955	-	-	3.11	1.33	1.35	.36	.08
1958	-	-	3.39	1.50	1.40	.36	.08

Source: U.S. Statistical Abstract, 1961, pp. 324-325.

2) Farm Households,[1] which also include agricultural corporations.
3) State and Local Governments
4) The Federal Government

Two intermediate sectors are distinguished: 1) unincorporated business and 2) corporate business. The distinguishing characteristic of an intermediate sector is that its equities are carried as assets of the non-farm household sector, and hence are regarded as liabilities of the intermediate sector, which therefore has zero net worth. When the wealth position of the intermediate sector is considered separately, the equity liabilities are ignored. Non-financial corporations are also shown as a sub-intermediate sector of corporate business. The detailed composition of all sectors is the same as that found in R. Goldsmith, Vol. III, from which the basic data are drawn.[2]

The income concept employed in all non-government sectors is that of income after taxes. In the government sectors it is that of total tax and non-tax receipts. The distribution of this income among the final sectors, given in Table 4, provides the system of weights which permits the reconciliation of the sectoral wealth-income ratios with the aggregate ratio. The system of weighting presented in Table 4 is not fully consistent with the income concept presented in calculating the aggregate ratios, i.e. NNP. The accuracy of our accounting identity was allowed to suffer in order to preserve the behaviorally significant concept of income.[3] However, our accounting identity did not suffer too greatly because the adjustments approximately offset one another. This is shown in line 11 of Table 4, where the total income of the final sectors is given as a percentage of NNP; in every year except 1933, intersectoral total income in Table 4 was a good approximation of NNP.

In Tables 5 and 6 the aggregate and sectoral wealth-income (or net worth-income) ratios are given with their two principal components, the gross asset-income and liability-income ratios. It is necessary to introduce a valuation adjustment in order to reconcile the capital-output ratio

---

1. All private corporate securities are carried in the non-farm household sector and the income from these securities is credited to that sector even though the ownership of U.S. Government securities by individuals is distributed between farm and non-farm households. This leads to an overstatement of less than 2 per cent of the wealth coefficient of farm households, and an even smaller understatement of the wealth coefficient of non-farm households.

2. Only post-1929 data are presented because of the difficulty in making the pre-1929 data comparable. The estimates shown for 1922 are given only to indicate an approximate order of magnitude.

3. A full reconciliation between the income concept utilized in Table 4 and NNP is given in R. Thorn, Appendix 3.

TABLE 4

Distribution of Income among Final Sectors

	1900	1912	1922	1929	1933	1939	1945	1949
	A. Absolute Values (Billions of current $)							
1. Non-Farm DI	13.0	26.0	55.5	77.2	43.4	66.2	139.3	176.3
2. Farm DI	2.0	4.1	3.8	5.9	2.4	4.3	11.1	11.9
3. Federal Government Receipts	.6	.7	3.4	3.8	2.7	6.7	42.5	39.1
4. State and Local Government Receipts	1.0	1.8	4.8	7.6	7.2	9.6	11.6	19.6
5. Total	16.6	32.6	67.5	94.5	55.7	86.8	204.5	246.9
	B. Per Cent of Total							
6. Non-Farm DI	78	80	82	82	78	76	68	71
7. Farm DI	12	13	6	6	4	5	5	5
8. Federal Government Receipts	4	2	5	4	5	8	21	16
9. State and Local Government Receipts	6	6	7	8	13	11	6	8
10. Total[1]	100	100	100	100	100	100	100	100
11. Line 5 as per cent of NNP	97	99	100	99	115	104	101	104

1. Parts may not add to total due to rounding.

Source: Various appendices in R. Thorn except line 3, 1900–1922, in Goldsmith, I, Table F-10, col. 11; and line 4, 1900–1922, in Goldsmith, I, Table G-1, col. 1, plus Table G-13, col. 1.

(where capital is valued on a replacement cost basis) with the wealth-income ratio (where corporate equities are valued at their market price).

In the constant price series, tangible assets with the exception of land were deflated by specific deflators, and intangible assets and income were deflated by the GNP implicit deflators. The rationale behind deflating intangibles by a general price deflator and not a specific deflator was that an individual values his intangible wealth in terms of the tangible goods it will buy. If the price of intangibles rises, the owner feels wealthier, provided there is no simultaneous rise in general prices.

There are some disadvantages attached to the form of the wealth data used as a basis for these calculations. Whereas R. Goldsmith gives annual estimates of the tangible national wealth, he gives estimates of intangibles only at selected dates. This deficiency is not particularly serious with regard to the absolute value of wealth since the stock of wealth changes slowly, but at times it does significantly affect the wealth-income ratio because of the greater variability of the denominator. This defect could have been partially overcome by averaging income over a period of years, but the wealth-income ratios would then be significantly influenced by the method of averaging chosen. In spite of some obvious shortcomings, the use of the annual ratios on selected dates was the method selected.

1. The Pre-1929 Adjustment

a. In Constant Prices

The adjustment that took place in the wealth coefficients of the various sectors in terms of constant prices will be examined first. As discussed above, the decline in the aggregate capital-output ratio before 1929 may be attributed principally to a decline in the ratio of land to output.[1] The corresponding financial adjustment of the wealth-income ratio was for the greater part intersectoral in nature. Agriculture, the sector which historically has had the highest wealth coefficient, experienced a fall in its share of national product from 13 per cent in 1912 to 6 per cent in 1922 (Table 4). The rate of decline was more gradual than these figures indicate, however, since 1912 was a boom year for agriculture and 1922 was a year of recovery from the agricultural depression of 1921. This intersectoral change in the distribution of income accounted for somewhat less than one-third of the total decline in the real wealth-income ratio before 1929.[2] The remainder of the decline was caused by a fall in the wealth coefficient of all the final sectors. The major

---

1. For convenience we speak of the adjustment to the change in the capital-output ratio. In general, one cannot regard the wealth-income or the capital-output ratio as being the independent member since both sides of our basic identity adjust mutually to one another. In specific instances, however, it may be possible to identify the factor which generates the initial disturbance.

2. See R. Thorn, Appendix 24, for more detailed evidence.

TABLE 5

Summary of Selected Asset-Income Ratios in Constant Prices[1]

	1900	1912	1922	1929	1933	1939	1945	1949
A. Net Worth								
1. Aggregate (Replacement Cost Basis)	5.7	5.1	5.0	4.8	5.9	4.2	2.5	3.4
2. Value Adjustment	-.5	-.0	-.4	+.4	-.1	-.1	-.2	-.5
3. Aggregate (Corporate Equities at Market Value)	5.2	5.1	4.6	5.2	5.8	4.1	2.3	2.9
4. Non-Farm Households	4.6	4.6	4.5	5.2	6.3	4.8	4.0	4.0
5. Farm Households	11.5	8.6	14.0	9.5	17.8	11.2	8.6	11.0
6. State and Local Governments	4.2	5.5	4.2	4.0	3.5	3.0	4.4	3.5
7. Federal Government	-.5	-.5	-3.9	-2.3	-5.5	-5.7	-5.1	-4.9
8. Unincorporated Business	-	-	-	4.4	7.1	4.2	3.7	4.0
9. All Corporations	-	-	(19.1)	18.0	-	18.7	21.3	16.5
10. Non-Financial Corporations	-	-	(17.9)	15.0	103.1	13.3	16.4	12.6
B. Total Assets								
1. Non-Farm Households	5.0	5.0	4.9	5.9	7.2	5.4	4.3	4.4
2. Farm Households	13.2	10.5	18.8	12.2	22.6	13.8	9.4	12.2
3. State and Local Governments	6.2	7.9	6.4	6.2	6.3	5.1	5.8	4.6
4. Federal Government	3.4	4.8	3.9	3.0	4.9	4.0	2.0	2.4
5. Unincorporated Business	-	-	-	7.0	11.9	6.2	4.7	5.3
6. All Corporations	-	-	(44.8)	41.7	-	60.7	67.4	45.8
7. Non-Financial Corporations	-	-	(30.4)	25.1	194.6	24.0	25.6	19.0
C. Total Liabilities								
1. Non-Farm Households	.4	.4	.4	.7	.9	.6	.3	.4
2. Farm Households	1.7	1.9	4.7	2.6	4.8	2.5	.8	1.1
3. Unincorporated Business	-	-	-	2.5	4.8	2.0	1.0	1.3
4. State and Local Governments	2.1	2.4	2.2	2.2	2.7	2.1	1.4	1.1
5. Federal Government	3.9	4.2	7.8	5.3	10.4	9.7	7.1	7.3
6. All Corporations	-	-	(25.7)	23.7	-	42.1	46.1	29.3
7. Non-Financial Corporations	-	-	(12.4)	10.1	81.8	10.7	9.2	6.4

1. Sum of parts may not equal total because of rounding.

Source: R. Thorn, various appendices.

portion of these intrasectoral declines occurred because income reached a cyclical peak in 1929, at a level which was high relative to that of the other benchmark years.[1]

The non-cyclical component of the fall in the net worth-income ratio of state and local governments was a result of the decline in their holding of land relative to income. During the period before 1929, the saving ratio of this sector actually increased.[2] The fall in the net worth-income ratio of the Federal Government had little effect on either the aggregate wealth coefficient or any of the other final sectors before World War I because of its small weight in the economy. The rise in the net worth-income ratio of the Federal Government in 1912 was offset by a decline in its weighting. After World War I the Government became a net debtor to the other final sectors of the economy, particularly the non-farm household sector.

The Federal Government supplied wealth claims to the private sectors during World War I. In view of the stability of the producer capital-output ratio in the postwar period the private community was able to attain a higher real wealth coefficient than the supply of private claims would have allowed. However, while non-farm households increased their holdings of Government securities by 4.1 per cent of total assets, in current prices, between 1912 and 1922, they reduced their holdings of corporate securities by 6.1 per cent during the same period.[3]

In Chart 2 the wealth coefficients of all sectors are shown in constant and current prices. It can be seen that the wealth coefficient of non-farm households moved within a relatively narrow range before 1929. If account were taken, as is done below,[4] of the meteoric rise in security prices in 1929, the variation is even smaller.

While there are wide swings in the agricultural wealth coefficient as a result of the cyclical instability of agricultural income (Chart 2), there is no evidence of any secular trend in its wealth coefficient.

b. In Current Prices

In Table 6 and Chart 2, it may be seen that the wealth coefficients in current prices exhibit the same general movements as the constant price coefficients except that the range of variation is somewhat smaller for the current price series.

The fall in the aggregate wealth coefficient from 1900 to 1929, in terms of current prices, was caused almost exclusively by the decline in the importance of agriculture in the economy; the fall in the wealth coefficients of agriculture and state and local governments was only a secondary factor.[5]

1. Cf. R. Goldsmith, Vol. II, p. 84, for a comparison of the cyclical position of the other benchmark years.

2. Cf. R. Goldsmith, Vol. I, pp. 121–133.

3. See R. Thorn, Appendix 7, for the source of these figures.

4. See Table 7, p. 267.

5. See R. Thorn, Appendix 25.

Chart 2

WEALTH COEFFICIENTS OF VARIOUS SECTORS

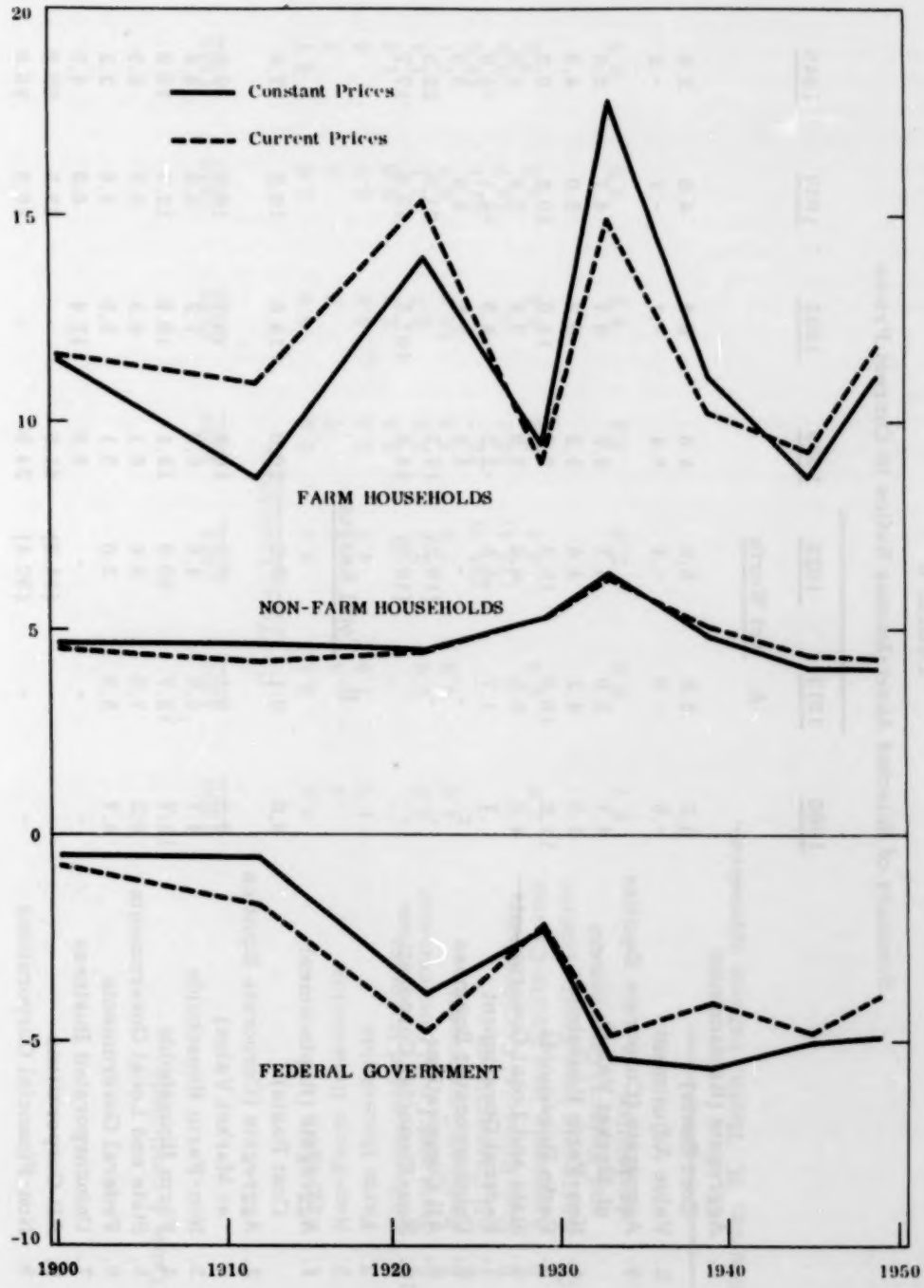


TABLE 6

Summary of Selected Asset-Income Ratios in Current Prices

	1900	1912	1922	1929	1933	1939	1945	1949
A. Net Worth								
1. Aggregate (Replacement Cost Basis)	5.2	5.0	5.0	4.6	6.8	4.8	2.8	3.8
2. Value Adjustment	-.5	-.0	-.4	+.4	-.1	-.1	-.2	-.5
3. Aggregate (Corporate Equities at Market Value)	4.7	5.0	4.7	5.0	6.7	4.7	2.6	3.3
4. Non-Farm Households	4.3	4.2	4.4	5.2	6.2	5.0	4.3	4.3
5. Farm Households	11.6	10.9	15.3	9.5	14.9	10.2	9.3	11.8
6. State and Local Governments	4.2	5.5	4.4	3.9	3.8	3.6	4.9	4.0
7. Federal Government	.7	1.7	-4.8	-2.2	-4.9	-4.1	-4.9	-4.0
8. Unincorporated Business	-	-	-	4.4	7.6	4.9	3.9	4.3
9. All Corporations	-	-	(19.2)	17.7	-	21.5	22.5	18.0
10. Non-Financial Corporations	-	-	(18.0)	14.8	107.5	15.6	17.7	14.2
B. Total Assets								
1. Aggregate (Replacement Cost Basis)	9.0	9.1	9.5	10.0	14.8	10.2	7.4	8.2
2. Aggregate (Corporate Equities at Market Value)	8.5	9.1	9.1	10.4	14.7	10.1	7.2	7.7
3. Non-Farm Households	4.7	4.6	4.8	5.9	7.2	5.6	4.6	4.7
4. Farm Households	12.7	12.7	20.0	12.1	19.6	12.7	10.0	13.0
5. State and Local Governments	6.2	7.9	6.6	6.1	6.5	5.7	6.3	5.1
6. Federal Governments	4.7	5.9	3.0	3.1	5.5	5.6	2.2	2.4
7. Unincorporated Business	-	-	-	6.9	12.4	6.9	4.9	5.6
8. All Corporations	-	-	(44.9)	41.5	-	63.5	69.0	47.2
9. Non-Financial Corporations	-	-	(30.4)	24.9	-	26.3	26.9	20.6

(Table 6 continued)

	1900	1912	1922	1929	1933	1939	1945	1949
			C. Total Liabilities					
1. Aggregate	3.6	4.1	4.5	5.4	8.0	5.6	4.7	4.5
2. Non-Farm Households	.4	.4	.4	.7	.9	.6	.3	.4
3. Farm Households	1.7	1.9	4.7	2.6	4.8	2.5	.8	1.1
4. Unincorporated Business	-	-	-	2.5	4.8	2.0	1.0	1.3
5. State and Local Governments	2.0	2.4	2.1	2.2	2.7	2.1	1.4	1.1
6. Federal Government	4.0	4.2	7.8	5.3	10.4	9.7	7.1	7.3
7. All Corporations	-	-	(25.6)	23.7	-	42.1	46.5	29.2
8. Non-Financial Corporations	-	-	(12.4)	10.1	-	10.7	9.2	6.4
9. Aggregate less Equity Claims	2.8	2.9	3.3	3.4	5.8	4.3	3.8	3.7
10. Aggregate less Equity Claims and Federal Government Liabilities	2.7	2.8	2.9	3.2	4.2	3.5	2.3	2.5

Source: R. Thorn, various appendices.

2. Summary: 1900–1929

Comparing the constant and current price data, we can see that not only did the wealth coefficients of the non-farm household sector remain relatively stable before 1929, but also that they displayed no trend, although the coefficients of the agricultural sector varied widely. The wealth-income adjustment to the fall in the aggregate ratio took the form of a declining importance of agriculture, the sector which characteristically possesses the highest wealth coefficient.

In addition to this broad movement there were many minor adjustments. The impact of the fall in the capital coefficient on the private sectors during World War I was partially offset by the assumption of a negative net worth by the Federal Government. After the war, in 1922, as the capital coefficient rose, the stability of the non-farm household coefficient was maintained largely by the low market value placed on private securities. As the negative net worth of the Government declined numerically, the price of private securities rose, and this movement culminated in the boom of the late twenties.

The overall adjustment to the decline in the importance of agriculture was achieved comparatively smoothly. This adjustment was marked by a shift of resources from agricultural to non-agricultural sectors of the economy—a shift which manifested itself largely in differential rates of growth. This adjustment did not seriously affect the structure of financial relationships between the sectors since the individuals in the agricultural sector not only owned the equity claims to agricultural wealth but also controlled the physical capital. The asset structures of other sectors, therefore, were only slightly affected.

3. The Post-1929 Adjustment

After 1929 and the Great Depression, the aggregate capital coefficient in constant prices fell from 4.8 in 1929 to 3.4 in 1949. This fall may be attributed largely to the decline in the ratio of plant and equipment to output during World War II, [1] in contrast to the previous quarter-century, where the decline was caused primarily by the falling ratio of land to output. The corresponding sectoral adjustment to the decrease in the aggregate wealth coefficient was of a complex nature.

The wealth coefficient of non-farm households, in constant prices, fell from 5.2 in 1929 to 4.0 in 1949. Much of this decline reflects movements in the price of securities, caused by short-run expectations. In the long run, equity prices seem to have fluctuated about their replacement cost value. With respect to their replacement cost, equities were greatly overvalued in 1929 and correspondingly undervalued in 1949. [2]

1. The exact date of the beginning of the decline may have been somewhat earlier. Cf. footnote 1, p. 255.

2. The pessimism which prevailed in the security markets in 1949 after World War II was somewhat analogous to that which prevailed in 1922

TABLE 7

Wealth Coefficients of Non-Farm Households with Corporate Equities Valued at Their Replacement Cost Value

	1900	1912	1922	1929	1933	1939	1945	1949
1. Constant Prices	5.3	4.6	4.9	4.8	6.4	4.9	4.3	4.7
2. Excluding Residential Land	5.1	4.5	4.7	4.5	6.1	4.7	4.2	4.6
3. Current Prices	5.0	4.2	4.8	4.8	6.3	5.1	4.6	5.0
4. Excluding Residential Land	4.7	3.9	4.5	4.5	5.9	4.8	4.4	4.8

Source: R. Thorn, various appendices; and R. Goldsmith, Vol. III.

In Table 7 the wealth coefficient of non-farm households is presented with corporate equities valued at replacement cost. This correction is an attempt to reduce the effect of short-run expectations in the security market on the wealth coefficient. This does not mean, however, that all security price movements which deviate from the replacement cost value of the securities are necessarily caused by short-run expectations. On the contrary, it will be shown that some of these changes form part of the process of long-run adjustment. However, for the particular years shown in Table 7, the adjustment serves to counteract the effect of short-run fluctuations in the price of securities. These fluctuations resulted from short-run expectations which were soon revised following the years 1922, 1929, and 1948. Another striking aspect of the data is the stability of the wealth coefficients over considerably long periods of time, if we disregard any definitely extreme cyclical or historical disturbances, such as those of 1933 and 1946.

The most striking development, however, was the sharp increase in the negative net worth of the Federal Government during World War II. At the same time its receipts increased from 8 per cent of national product in 1939 to 16 per cent in 1949 (Table 4). On one hand, the Federal Government absorbed a larger portion of the national product and therefore tended, ceteris paribus, to raise the wealth coefficient of the private sectors by reducing income after taxes. On the other hand, it provided these sectors with claims to wealth, in the form of U.S. government securities, which did not represent additions to the productive capacity of the economy.

after World War I, and the undervaluation relative to replacement cost was roughly of the same magnitude. Cf. Table 5.

The increase in indebtedness of the federal government coincided with the fall in the capital coefficient and enabled the private community to hold a greater amount of wealth per unit of income than would otherwise have been possible, ceteris paribus. The ceteris paribus excludes, in particular, the possibility of a rise in the consumer capital coefficient to offset the fall in the producer capital coefficient. It is highly doubtful that such an adjustment would have been possible, in any case, since it would have been necessary for consumers to hold almost twice as much consumer capital per unit of income as they had held in any period covered by the study. Even if consumers had wanted to hold an additional amount of consumer durables per unit of income, the consumer goods industries would have been incapable of sufficient output in the short run. As it was, there were still some shortages of consumer goods in 1949.[1]

The ratio of U.S. Government securities to the disposable income of non-farm households was .97 in 1949 as compared with .12 in 1939. The difference between these two figures provides a rough measure of the influence of the Federal Government in enabling the non-farm sectors to attain a higher wealth coefficient than would otherwise have been possible without sweeping changes in the structure of private assets. The private community could not attain a higher wealth coefficient simply by increasing its investment in private industry in that this investment in capital at the existing producer capital-output ratio would have generated a quantity of additional income sufficient to leave the wealth coefficient unchanged.[2]

The Federal Government, by supplying wealth claims to the private community during a period when the producer capital coefficient was declining rapidly, prevented the full force of adjustment to a lower aggregate wealth coefficient from falling on the private sectors of the economy.

---

1. It is interesting to note that the wealth-income ratio, and more particularly its consumer durable component, gives some objective measure of what is commonly called "backlog demand." For example, if this data had been available in 1949 it would have been evident that non-farm households had attained a consumer durable income ratio that was close to its historically stable value and that, unless some strong structural changes in the nature of consumer preference had occurred during the war years, the "backlog demand" had been almost eliminated. This, however, must be treated as only a rough measure since many factors affect the various wealth asset ratios, some of which will be discussed below in our model of adjustment. See R. Thorn, Appendices 3 and 4.

2. This assumes there is no time lag between investment and the increase in output. If lags are introduced, a higher wealth coefficient might be attained for a short period (under certain dynamic assumptions, however, the private community might permanently achieve a higher wealth coefficient).

However, this partial adjustment will be permanent only if the federal government's indebtedness continues to grow at the same rate as national income.

4. Summary of the Historical Data

A review of the empirical data for the entire half-century, in spite of its shortcomings, leads to certain general conclusions. The capital-output ratio in real terms fell over the first quarter-century, primarily as a result of the stock of land becoming proportionately smaller as the stock of other assets grew. During World War II, or possibly earlier, a significant fall occurred in the producer component of the capital-output ratio, and this further contributed to the decline in the wealth-income ratio caused by the downward trend of the land-output ratio. The fall in the producer component of the capital coefficient, in real terms, continued into the postwar period and, while the producer component may possibly have a rising trend at present, in 1955 it was still well below its stable prewar level.

The fall in the aggregate wealth coefficient that corresponds to the decline in the capital coefficient is largely attributable to the diminishing importance of the agricultural sector. In the twenties this secular reduction in the income share of agriculture, combined with intrasectoral declines in the wealth coefficients of agriculture and of the federal government, was sufficiently strong to offset the rise in net worth of the non-farm sector caused by the stock market boom.

During World War II the aggregate wealth coefficient fell sharply but, because of the sharp rise in the negative net worth of the Federal Government, the decline was less severe in the private sectors of the economy. The Federal Government supplied wealth claims to the private sector which did not correspond to increases in capital stock which would generate additional income.

Throughout the entire period under review, the wealth coefficient of the non-farm sector showed a high degree of stability in constant prices, and an even higher degree in current prices. These statements apply *a fortiori* when adjustment is made for short-run movements in security prices.

This brings us to the problem of adjustment now (1957) facing the economy. The private wealth coefficients have historically fluctuated within a very narrow range. The producer component of the capital-output ratio in real terms has fallen sharply and there is little reason to expect the national debt to grow proportionately to national product in the foreseeable future. Therefore, either the private sectors of the economy must accept a much lower wealth coefficient in real terms than they have historically held or they must increase their holdings of consumer capital per unit of income—a process which requires a higher rate of saving through consumer durables. However, if the capital coefficient of producers rises, the extent of the adjustment required of the private wealth-holding sectors will diminish. In the next section a simple model is introduced to demonstrate some of the possible mechanisms of adjustment.

E. The Model of Adjustment

The model of adjustment which will now be presented calls attention

to the relationship between the essentially financial decision to accumulate wealth and undertake saving, and the technological-financial decision to employ a given size capital stock per unit of output. It emphasizes the role of capital intensity in production as part of the adjustment mechanism, as well as changes in the price level and employment. The form of the model is a variation of one presented by Tobin,[1] but the somewhat different treatment accorded the role of government securities and the more detailed examination of the effect of the substitution between capital and labor lead to some new results.

The model has two wealth supplying sectors: the private productive sector and the government sector. The private productive sector holds the physical productive capital of the economy.

In the short period the model contains five exogenous variables and twelve endogenous variables, and consists of twelve equations (5-16). The variables are:

Exogenous Variables

- K = stock of physical producer capital
- N = rate of input of labor services
- G = number of government securities
- v = coupon rate of interest on government securities
- p = general price level of consumer goods

Endogenous Variables

- Y = real rate of output at any point in time
- W = stock of "real" wealth of the private sector
- D = stock of physical consumer capital
- M = stock of non-coupon bearing government debt, i.e. money
- w = money wage rate
- r_K = rate of return on producer capital
- r_C = rate of return on claims to producer capital
- r_G = rate of return on the market value of coupon bearing government debt
- P_K = price of physical producer capital
- P_C = price of claims to physical producer capital
- P_D = price of physical consumer capital
- P_G = price of coupon bearing government debt

1. The Supply of Wealth

The supply of wealth consists of claims against producer capital, coupon bearing claims against the government, non-coupon bearing claims

1. J. Tobin, "A Dynamic Aggregative Model," Journal of Political Economy, Vol. LXIII, No. 2 (April 1955), pp. 102-115.

against the government, i.e. money,[1] and consumer capital. The demand for producer capital is the supply of wealth claims to producer capital.

a. The Production Function

Output is produced at the rate (Y) at any point in time and is equal to the income of the community. The rate of output with a given technology depends on both the stock of producer capital and the rate of input of labor services:

$$(5)\quad Y = \psi(K,N)$$

Output adjusts simultaneously to changes in factor inputs so that no lag exists between increases in factor inputs and increases in output. The supply of labor services is determined exogenously.

All production functions determined by the given technology are linear homogeneous so that the marginal products of labor and capital depend only on the proportion in which the two inputs are being used. The real wage rate and the rate of return on physical producer capital are equated to their marginal physical products by competition.[2]

$$(6)\quad \frac{w}{P} = \psi_N(K,N)$$

$$(7)\quad r_K = \psi_K(K,N)$$

One of the properties of a linear homogeneous production function, though not unique to it, is that the producer capital coefficient (K/Y) varies inversely with changes in the rate of return on physical capital (r_K).[3]

Initially it will be assumed that expectations remain static, in that the prices and technology which currently prevail are expected to continue into the future. The marginal productivity of capital may therefore be assumed equal to its marginal efficiency.[4]

b. The Supply of Government Claims

The number of coupon bearing government claims and the amount of non-coupon bearing government claims can be changed only by

---

1. At the present level of aggregation, demand deposits are not considered as money since depositors and bankers are both members of the private sector and their claims offset one another. As before, consumer capital is defined as consumer durables and residential housing.

2. The notation, $\psi_N(K,N) = \frac{\partial\psi(K,N)}{\partial N}$, etc.

3. The simple proof of this, as well as a lengthy discussion of the relationship between the marginal productivity of capital and the average productivity, is given by F. M. Bator, "On Capital Productivity, Input Allocation and Growth," Quarterly Journal of Economics, LXXI, No. 1 (February 1957), pp.86-106. The proof referred to is given on page 92.

4. Cf. A. P. Lerner, The Economics of Control (New York: MacMillan, 1944), pp. 330-335.

government budget deficits and surpluses. The number of government coupon bearing claims is determined exogenously while the amount of non-coupon bearing claims supplied by the government always equals the amount the private sector wishes to hold and is determined endogenously. The general price level of consumer goods is assumed to remain unchanged. A more realistic way, perhaps, of describing the behavior of the money supply would be to say that the government attempts to vary the supply in order to maintain a stable price level.

c. Interest Rate Relationships

All coupon bearing government securities are issued at a fixed par value with the same coupon rate of return (v), so that changes in the rate of interest on these securities are related to changes in their price:

$$(8)\quad r_G = \frac{v}{p_G}$$

The price of government securities is determined in a free market and is always such that the demand for government securities equals the number supplied.

The rate of return on claims to producer capital is related to the rate of return on physical producer capital by the ratio of the price of producer capital to the price of claims on producer capital:

$$(9)\quad r_C = \frac{P_K}{P_C} \cdot r_K$$

It is assumed that the price of producer capital will vary with the general price level so that:

$$(10)\quad P_K = p$$

We shall comment on this assumption later in the analysis.

2. The Demand for Wealth

The private sector measures its wealth assets in terms of the amount of consumer goods for which these assets may be exchanged.[1] The "real" wealth value of an asset to the private sector is its market value, deflated by the general consumer price index. The effect of price changes on the wealth value of any asset may be measured by the ratio of its market price to the index of consumer prices. For the sake of simplicity, it is assumed that the price of consumer durables will move with consumer prices in general, so that:

$$(11)\quad P_D = p$$

---

1. See above, p. 262.

The wealth coefficient of the private sector (W/Y) is defined by the following identity:

$$(12)\quad \frac{W}{Y} = \frac{P_c \cdot K}{pY} + \frac{P_D D}{pY} + \frac{P_G G}{pY} + \frac{M}{Y}$$

The terms of (12) represent the wealth portfolio of the private community.

a. Wealth-Holding Behavior

The basic hypothesis concerning the aggregate behavior of the private community with respect to the holding of wealth is that the amount of wealth demanded per unit of income is a function of income[1] $\left[\frac{W}{Y} = f(Y)\right]$, but that the particular portfolio of assets depends on the relative rates of return received from each asset. For simplicity of exposition, it will be assumed, not wholly unrealistically, that the desired wealth coefficient of the private sector is a constant:[2]

$$(13)\quad \frac{W}{Y} = \alpha$$

b. Asset Preferences

The composition of the wealth portfolio desired by the private sector is determined by the rate of return on claims to producer capital, the rate of return on coupon bearing government debt, the price of consumer durables relative to other consumer prices (the real rate of return on physical consumer capital being assumed constant), and the own rate of return on money. The own rate of return on money shall be assumed to be zero. The ratio of the price of consumer durables to the index of consumer prices has

---

1. If some additional assumptions can be made, this hypothesis is consistent with some recent consumption theories put forth by M. Friedman, F. Modigliani and R. Brumberg. The most important of these assumptions is that the population has a constant age distribution and that there is homogeneity within the age groups. Further, it is assumed either that the income and substitution effects of the rate of interest counteract one another or that their net effect is negligible. There is some empirical evidence to support this latter assumption, since most savings studies have found the rate of interest a relatively unimportant determinant of savings behavior. Cf. M. Friedman, A Theory of the Consumption Function; F. Modigliani and R. Brumberg, "Utility Analysis and the Consumption Function" in Post-Keynesian Economics, K. K. Kurihara, ed. (New Brunswick: 1954); and R. Goldsmith, Vol. III, Part 4.

2. Implicit in (13) is the assumption that the elasticity of substitution with respect to the variables of the model is absolutely greater among the various forms of wealth than between wealth assets and consumption goods. Cf. I. Friend and V. Natrella, Individuals' Saving: Volume and Composition (New York: Wiley, 1954), p. 14.

already been assumed constant (11); therefore the asset preferences of the private sector depend only on the rates of return on claims to producer capital and coupon bearing government debt:

$$(14)\quad \frac{P_C K}{pY} = h\,(r_C,\ r_G)\ ;\ h_{r_C} > 0\ ,\ h_{r_G} < 0$$

$$(15)\quad P_D D = j\,(r_C,\ r_G)\ ;\ j_{r_C} < 0\ ,\ j_{r_G} < 0$$

$$(16)\quad P_G G = g\,(r_C,\ r_G)\ ;\ g_{r_C} < 0\ ,\ g_{r_G} > 0$$

The equation for the amount of money held per unit of income is omitted, the aggregate wealth coefficient being determined by (13): once three of the four types of wealth assets are determined, the fourth is determined also.[1]

It will be assumed, realistically, that the absolute value of the cross elasticity of demand for government bonds with respect to changes in the rate of interest on claims to producer capital is less than the elasticity with respect to their own rate of interest; and/or that an analogous relationship exists between the demand for producer capital and the rate of interest on government bonds, i.e.

$$\eta_{g_{r_G}} > \left|\eta_{g_{C}}\right| \quad \text{and/or} \quad \eta_{h_{r_C}} > \left|\eta_{h_{r_G}}\right| .\ ^{2}$$

It is also assumed that, at higher rates of interest on financial assets, less wealth will be held in the form of consumer durables and more in the form of financial assets. The rationale behind this assumption is that the flow of services from a sum invested in consumer durables will appear relatively smaller in terms of its money equivalent when compared with the rate of return that could be attained by investing the same sum in financial assets. It would be relatively cheaper to obtain the services of consumer durables, if they were available for sale, by purchasing them with the yield of a financial asset than by buying the consumer durable directly.

If, at the existing interest rates and prices, the private sector is willing to hold the given amount of K and G, and also the existing stock of D and M, then the situation is one of "portfolio balance"; also, if the income distribution relations (6) and (7) are satisfied, the situation is one of short period equilibrium. If the community is not willing to hold the given assets at the existing prices and interest rates, then it will bid up the prices of those assets

---

1. If one of the other asset equations were omitted, we could substitute the equation for money: (16a) $\frac{M}{pY} = 1\,(r_C,\ r_G)\ ;\ 1_{r_C} < 0,\ 1_{r_G} < 0$

2. The notation $\eta_{g_{r_G}}$ denotes the cross elasticity of demand of government bonds with respect to r_G, etc.

of which it desires to increase its holdings, forcing down the price of assets of which it desires to decrease its holdings. At some combination of prices and interest rates the private community will be satisfied with the given amount of K and G and the existing stocks of D and M.

c. The Long Period Model

In the long period, however, where the stock of producer capital is free to vary, the rate of return on claims to producer capital must equal the rate of return on physical producer capital:

$$(17)\quad r_C = r_K$$

This is equivalent to saying that the price of claims to capital must equal the price of capital in the long run, as can readily be seen from (9). If, for example, the price of producer capital were below the price of claims on producer capital, producers would always be willing to increase the size of their capital stock and sell the claims to the additional capital to the private sector. Similarly, if the price of claims were below the price of physical producer capital, producers could raise their rate of return by purchasing claims and taking possession of the physical capital they represent.

F. Equilibrium

The model achieves a stationary equilibrium in which "induced" saving and investment occur only in the process of attaining equilibrium. Once an equilibrium has been reached, a new disturbance must occur before additional saving can take place. The method of analysis employed will be that of comparative statics; starting at a point of equilibrium at full utilization of the capital stock and labor force, a specific disturbance is introduced and an analysis is then made of the mechanism of adjustment until a new point of equilibrium is attained.

The possible consequences of assumptions alternating to those presented in the model will be discussed in a less formal framework.

It is obvious from a consideration of the wealth-holding behavioral equations (12–16) that, if the stock of all assets, including government debt, changed in direct proportion to changes in income, "portfolio balance" would always be attained at the existing prices and interest rates. The disturbances that will be singled out for special analysis are those that are not proportional to changes in income, namely, changes in the stock of producer capital and the national debt.

G. Adjustments to Changes in the Producer Capital Coefficient

What happens when the producer capital coefficient (K/Y) changes without a simultaneous shift in private asset preferences so that, at the existing prices and interest rates, the stock of physical producer capital (K) is insufficient to satisfy the private community's demand for claims to producer capital as determined by (14)?

If the realized wealth coefficient differs from the desired wealth

coefficient (α), one possibility open to the private community in order to attain its desired wealth coefficient in a finite period of time is to increase its rate of saving.[1]

If the increase or decrease in the realized wealth coefficient necessary to restore the desired wealth coefficient is of sizeable magnitude, the attainment of the new ratio might require considerable time. The maximum rate of increase would occur if the community saved its entire income but, for obvious reasons, the maximum practical value is considerably below this. The maximum fall in the real wealth-income ratio, at constant prices, is set by the rate of depreciation of the community's tangible assets when the community saves none of its income. However, the community does have another and much quicker means of attaining a higher or lower wealth coefficient as a consequence of the method by which it was assumed the private community computes its real wealth. The market value of claims to wealth is deflated by a price index indicating the general purchasing power of money. By bidding up the price of intangible claims to wealth at a faster rate than the general price level, the private community can attain a higher wealth coefficient in what it considers "real terms." This may be called the "wealth effect" of price changes.[2] It represents an illusion in the sense that the private community feels wealthier (because of the rise in the price of intangible assets relative to the price of consumer goods) while there has been no change in the stock of real capital goods. In the mechanism of adjustment to changes in the producer capital coefficient, both movements in relative prices and variations in the rate of saving play an important role.

A change in the capital coefficient desired by producers may be brought about by a wave of producer capital-saving, or capital-using, innovations; by changes in the composition of output; or by substitution of capital

---

1. In a dynamic situation, not considered by the model, this would suggest the hypothesis that the saving ratio might be expressed as a function of the desired wealth coefficient and the realized wealth coefficient. If the desired wealth coefficient is assumed to remain constant, then the saving ratio is merely a function of the realized wealth coefficient. The formal relationship between a change in the aggregate wealth coefficient $\left(\frac{U}{Y}\right)$, the saving ratio $\left(\frac{dU}{Y}\right)$, and the rate of income growth $\left(\frac{dY}{Y}\right)$ is given the following identity:

$$(18)\quad d\left(\frac{U}{Y}\right) = \frac{dU}{Y} - \frac{U}{Y}\left(\frac{dY}{Y}\right)$$

In one empirical study a high negative correlation was found between the realized wealth coefficient and the saving-ratio. See W. D. Hamburger, "The Relation of the Consumption Function to Wealth and the Wage Rate," Econometrica, Vol. 23 (January 1955), pp. 1–17.

2. This "wealth effect" is, in fact, a type of generalized "Pigou effect."

for labor (caused by changes in their relative scarcities) along given production functions.

For purposes of illustration, the adjustment to a fall in the producer capital coefficient, caused by a wave of capital-saving innovation, will be analyzed within the framework of our model. Before proceeding it is worthwhile to define precisely a capital-saving innovation. Given a linear homogeneous production function, the stock of producer capital and the input of labor services, a capital-saving innovation is one that raises the marginal productivity of capital less than the average productivity so that the relative share of labor increases. It immediately follows that a capital-saving innovation raises the marginal product of capital less than output. [1]

A capital-saving innovation may also be defined as one such that, if producers were to increase their capital stock to the point where the marginal productivity of producer capital were equal to its marginal productivity before the innovation was introduced (the labor input being held constant), the average productivity of capital would be greater, i.e., if producers attempted to keep the marginal productivity of capital constant in the face of capital-saving innovations, the producer capital coefficient of production would fall. [2] Again the relative share of capital would decrease and the relative share of labor would rise.

The capital-saving innovations considered here will be considered capital-saving in both of the above senses, i.e., in both the Hicksian and Robinsonian formulations.

1. Adjustment in the Short Period

In the short period it is assumed that the producer capital stock remains unchanged. The effect of a capital-saving innovation is to increase output and the marginal product of producer capital relatively more than output.

The rise in output causes a desire on the part of the community to increase its holdings of wealth and all its components in proportion to the increase in income at the existing prices and interest rates. However, the quantity of producer capital and government bonds is fixed by assumption. The private community, in their attempt to acquire more claims to producer capital and government bonds, will bid up the price of these assets, causing the ratios $\frac{P_C}{p}$ and $\frac{P_G}{p}$ to rise, increasing the wealth value of these assets and, at the same time, causing r_C and r_G to fall. [3] The fall in r_C and r_G will cause substitution in favor of consumer durables and money.

1. Cf. J. R. Hicks, The Theory of Wages (London: MacMillan, 1932), pp. 121–130.

2. Cf. J. Robinson, "Classification of Inventions," Review of Economic Studies, Vol. 5 (February 1938), pp. 139–142.

3. For r_C to fall, it is necessary that P_C rise faster than r_K, as can be

In the short period, therefore, before changes in r_C can affect the capital coefficient of producers, a temporary equilibrium will be established through a rise in the price of intangible claims to wealth, relative to consumer prices in general, and through substitution effects induced by the fall in the rates of interest. The long-run equilibrium condition (17) is not satisfied.

2. Adjustment in the Long Period

In the long period the size of the stock of producer capital is free to vary. The lower rate of interest on claims to producer capital will induce producers to increase the size of their capital stock. This increase will cause the rate of return on producer capital (r_K) to fall and the producer capital coefficient (K/Y) to rise. Producers will continue to add to their capital stock until, for reasons which have already been discussed, $r_K = r_C$.

The increase in the amount of producer capital supplied per unit of income will cause a fall in the price of claims to producer capital and a rise in r_C. However, since the new investment increases output while the supply of government bonds remains constant, r_G will fall further. The only question left is whether r_C will remain below its initial value before the innovation was introduced. In long-period equilibrium it is known that $P_C = P_K$ from (9) and (11) and that $P_C = p$ from (10), so that the ratio P_C/p must equal one in (14). From the definition of a capital-saving innovation it is known that for long-period equilibrium, if r_C is equal to its initial value before the innovation was introduced, then K/Y must be below its initial value since $r_C = r_K$ in long-period equilibrium. If K/Y were below its initial value, (14) could be satisfied if r_G were to rise above its initial value. However, it has already been shown that r_G must fall, so that the only way (14) may be satisfied is if r_C falls below its initial value.

Since we know that r_C and r_G must decline relative to their initial value, it follows that substitution will take place in favor of D and M in the long period as well as in the short period.

3. Summary of the Adjustment Process

The adjustment in both the short and the long period, when compared with the initial point of equilibrium, is achieved through a fall in the rates of interest and through substitution in favor of consumer capital and money. In

---

seen from (9). This can easily be shown to be true if r_G remains constant. If r_C were to remain constant, P_C would have to increase by the same percentage as r_K from (9). However, the percentage increase in r_K is less than that in Y by the definition of a capital-saving innovation. Hence P_C would also have to increase less, percentagewise, than Y. However, if P_C increased by a smaller percentage than Y, while r_C remained unchanged, (14) would not be satisfied. The only way (14) could be satisfied is if P_C rose and r_C fell. The proof of the case where r_G is also free to vary is given in R. Thorn, Mathematical Appendix.

the short period, however, the major burden of adjustment falls on changes in interest rates and prices while, in the long period, some of these initial movements are partially reversed when the stock of producer capital is free to vary.

A similar analysis would apply even with production functions where the producer capital coefficient were inelastic with respect to changes in r_C, provided the marginal productivity of producer capital falls as the stock of producer capital increases.[1] However, for long-period equilibrium, either the producer capital coefficient or the demand of the private community for claims to producer capital per unit of income must be elastic with respect to r_C ($r_C = r_K$).

The consequences of violating various assumptions of the model are discussed below in a less formal framework than has been used so far.

4. The Possibility of a "Capital Shortage"

It is possible that innovations may create a demand for capital goods that is beyond the capacity of the capital goods industries to satisfy. There are some practical reasons why entrepreneurs should not immediately increase the rate of expansion of the capital goods industries. They have no way of knowing whether the wave of innovations will be sustained, and they may fear the possible operation of an "acceleration principle" if it should not be sustained. This view may be justified for a long period of time because of the wide fluctuations which normally occur in the demand for investment goods. Under these circumstances, even if a long-run rate of interest on producer capital is established which would equate producers' desired capital coefficients in the long run to the private community's desire to attain a given wealth coefficient, it is possible that a position of equilibrium may not be achieved because of a "capital shortage." Even if it were possible for the Government to increase the rate of interest alone, this measure would be ineffectual unless the demand for capital claims by the private community were relatively interest inelastic and the demand for capital goods by producers were relatively interest elastic. In fact, the opposite situation seems more likely. In the short run, a rise in the rate of interest (while it might cause a slight reduction in the rate of investment) would increase the availability of capital funds by means of the substitution of saving through claims to producer capital for saving through consumer durables and money; as a result, any producer who desired to obtain funds at a lower rate of interest would find willing lenders. The only way that a higher rate of interest could cause a reduction in the rate of investment would be for the government to supply the private community with "equivalent claims" to wealth,[2] thus

1. Cf. F. M. Bator, op. cit., for examples of such production functions.

2. "Equivalent government claims" would be those of which the expected yield would be considered equal to that of claims to producer capital so that an investor would be indifferent to which of the two he held. In the real world, with a variety of securities which include equity claims and holders of securities who are risk-seekers this formulation would require modification.

shifting the demand for private claims to capital per unit of income. The long-run problem, if the rate of innovation is expected to continue, and the short-run problem, if it is not, is how to achieve an increased expansion of the capital goods industries so that the economy may attain the higher rate of growth implied in the higher marginal productivity of capital.

This is the same problem that D. H. Robertson tackled some thirty years ago.[1] To achieve the necessary expansion, capital goods prices must rise relatively to consumer goods prices. If the general price level is to be stabilized, consumer goods prices must fall, capital goods prices must rise, and the money supply must increase. If consumer prices are stabilized, capital goods prices must rise even higher and the increase in the money supply must be greater.

It is unlikely that the expansion will be achieved very efficiently, as D. H. Robertson pointed out, when there is a relative rise in the price of capital goods but a fall in absolute prices, since this would require, among other things, that wage rates be very flexible in a downward direction.

Assuming prices are flexible, the "capital shortage" just described will cause capital goods prices to rise relative to consumer goods prices and, if the rise is substantial enough and producers have favorable expectations, the necessary expansion will take place. The increase in capital goods prices, however, will lower the marginal net revenue product of producer capital and cause producers to substitute labor for capital, thus decreasing the desired capital stock and lowering the capital coefficient employed at given rates of interest. The size of the wage increase that would follow depends on the elasticity of the supply of labor, which has not been considered. As the "capital shortage" is relieved, the relative price of capital and the rate of interest should fall. The wage rate might, or might not, fall, depending on the increase in the marginal product of labor, if any, that took place during the period of adjustment. The movements in nominal prices would, of course, depend on the changes in the money supply which took place. These cases were analyzed by D. H. Robertson and need not be repeated here.

5. A Rise in Wage Rates in Excess of the Increase in Labor Productivity

If the wage rate should rise in excess of the marginal product of labor, then the effects of the shift in the capital coefficient schedule, caused by the increase in capital goods prices as a result of the "capital shortage," would be partially offset or perhaps even reversed.

If the role of expectations is interjected for a moment into the analysis, the marginal efficiency of capital and the marginal productivity may

---

1. D. H. Robertson, Banking Policy and the Price Level (New York: Augustus M. Kelley, 1949 (reprint)), pp. 29–33. What follows is merely a summary of D. H. Robertson's conclusions.

differ. If producers expect wage rates to rise, they will tend to utilize a somewhat higher capital coefficient, when making a fixed investment of considerable durability, than the existing price of capital and labor indicate. This is because a slightly more capital intensive investment partially protects the producer from future wage rate increases.

6. Substitution Effects

To achieve the necessary expansion of the capital goods industries, the price of these goods must rise relatively to consumer goods. It is assumed that deflation in consumer goods prices is ruled out as a means of adjustment because of the downward inflexibility of wage rates, which might impose an undesirable level of unemployment on the community. The next best policy would seem to be for the monetary authorities to permit an expansion in the money supply sufficient to allow capital goods prices to rise while maintaining the level of consumer goods prices stable.[1] Can such a policy achieve success? The theoretical condition for the success of such a policy is that the marginal rate of substitution between capital goods and consumer goods to their users be zero. The more easily consumer goods can be substituted for capital goods, the quicker the mechanism of adjustment will proceed, but the consumer price level will be forced higher.

In the real world the rate of substitution of capital goods for consumer goods is not zero, and this is particularly true of primary products. Steel is a capital good when it goes into a machine tool and a consumer good when it goes into a bathtub.

If the increased rate of innovation is only temporary, the price of capital goods will fall again; however, the general price level may be somewhat higher, because of the monetary expansion, if the rise in output has not increased the demand for money beyond the additional supply. If the increased rate of innovation is permanent, equilibrium may be obtained; capital goods prices will be stable at a higher level and the rate of increase in the money supply will be able to remain constant.

The maximum rate of income growth with a given stock of innovations may not be the primary consideration. Instead, for well-known reasons, price stability may be more desirable from a welfare point of view. In this case, if wage rates are sticky in a downward direction, the alternative adjustment, in which the rate of growth of the money supply is kept constant, would achieve a stable price level at the cost of some unemployment in the consumer goods industries.

H. Adjustment to Changes in the National Debt

Returning to the formal model, the adjustment to a change in the amount of coupon-bearing government debt, i.e. government bonds, will

1. This was approximated by D. H. Robertson's change in the money supply to reflect changes in the real demand for machines. See D. H. Robertson, Banking Policy and the Price Level, pp. 29-33.

be considered. As before, the economy is assumed to be in long-period equilibrium at full employment and full utilization of the capital stock. The disturbance is introduced by assuming a fixed percentage increase in the input of labor services (N) and the stock of producer capital (K). If the stocks of all other assets were free to vary, they would increase by the same fixed percentage so that prices and interest rates would remain unchanged; it shall be assumed, however, that the number of government bonds (G) remains constant. As the initial point of departure it will be assumed that the stocks of all assets except government bonds have increased by the same percentage as income, and the adjustment process will be traced from that point.

If, at a given structure of interest rates and prices, the government decided to redeem all its bonds by exchanging them for non-coupon bearing government debt, i.e. money, then all government security holders could be divided into two groups—those who initially would be willing to hold the money given them, and those who would seek to purchase another earning asset as a substitute for the government securities. Those who are willing to hold money under these circumstances are holding government securities merely as a very close money substitute which yields a small income; they will be called liquidity holders of government securities. The other group consists of those who desire a very safe form of investment: they will be called investment holders of government securities. The possible paths of adjustment depend on the relative sizes of these two groups and on the government's willingness to supply as much non-coupon bearing debt as the community desires to hold.

If the possibility is allowed that all holders of government bonds may be either liquidity holders or investment holders, it is also necessary that one member of the pair of partial derivatives specified in (14), (15), (16), and (16a) be allowed to become zero, reflecting the possibilities of zero cross-elasticities of demand.

1. Adjustment with the Money Supply Free to Vary

In the short period the list of exogenous and endogenous variables are the same as those employed in the analysis of capital-saving innovations. The supply of money is assumed free to vary, the price level and the stock of producer capital remaining constant.

It follows from the description of the initial disturbance that, in the short period, the wealth holding equation (13) and the government bond preference equation (16) are not satisfied. It also follows from the initial assumptions that (13) can only be satisfied by a rise in at least one of the variables P_C, P_G, D, and M, if we exclude the possibility of a fall in Y.

a. Effect of Liquidity Holders

If all would-be government bond holders, at the initial prices and interest rates, are liquidity holders, then the adjustment takes the form of a fall in r_G as they bid up the price of government bonds. By assumption, the liquidity holders do not regard private capital claims as substitutes for government securities, so P_C and D remain unchanged. The form in

which the private community satisfies its demand for increased liquidity has merely changed from government bonds and money to money alone.

b. Effect of Investment Holders

If all would-be government security holders are investment holders, at the initial prices and interest rates, then the adjustment will take the form of a rise in P_G and a rise in P_C as they bid up the price of government bonds and substitute private claims to producer capital. The rises in P_G and P_C cause r_G and r_C to fall, and the decline in r_C will induce substitution in favor of consumer capital and money. The new equilibrium has been achieved through a fall in r_G and r_C and a rise in M and D.

In the long period, when the stock of producer capital is free to vary, producers will tend to increase their capital coefficient and their stock of capital, thus causing r_K and P_C to fall. The final value of r_C will be below its initial value before the short-period adjustment took place. This follows from (9) and (12) (remembering that it may be shown that, in long-period equilibrium, $P_C = p$),[1] since it is known that as K increases r_K must fall. The long-period adjustment is thus similar to the short-period adjustment when both are compared with the initial equilibrium. The rates of interest, r_C and r_G, will fall and there will be substitution in favor of money and consumer durables.

c. Mixed Case

The case where both liquidity holders and investment holders of government bonds are unsatisfied presents no special problem. M increases and r_G falls proportionately to the number of would-be liquidity holders of government securities. The remainder of the adjustment will be the same as described in the previous section, so that the net result will still be a fall in the rates of interest and substitution in favor of M and D.

2. Adjustment with a Constant Money Supply

The original assumptions of the model will now be changed slightly. The supply of money will be assumed to be exogenously determined by the government and the price level will be free to vary. It will also be assumed that the money supply will be increased by the same percentage as K, N, and Y in the introduction of the initial disturbance. The number of government bonds remains unchanged. The only way the wealth holding equation (13) may be satisfied in the short period is by changes in p, P_C, P_G, and D.

a. Effect of Liquidity Holders

In the short period, if all the would-be holders of government bonds, at the existing prices and interest rates, are liquidity holders, they will bid up the price of government bonds, causing r_G to fall and thus increasing the demand for money. The fall in r_G will not affect the demand for private capital claims since, by assumption, the liquidity holders do not regard

1. See above, p. 275.

private capital claims as adequate substitutes for government securities.[1] The only way the private community may increase its real cash balances is through a fall in p, the general consumer price level. As p falls, it follows from the initial assumptions that the private community will find that the amount of claims to producer capital that it holds is in excess of the amount it desires (14), with the result that P_C falls and r_C rises. The rise in r_C will cause substitution of claims to producer capital for consumer durables. A new short-period equilibrium is thus attained through a fall in r_G, p, P_C, and D, and a rise in r_C and P_G.

In the long period, when the stock of producer capital is free to vary, the rise in r_C will induce producers to employ a lower capital coefficient and a smaller capital stock, which will result in a decline in output since the labor input remains constant.[2] This will result, in turn, in a partial reversal of the initial price and interest rate movements; however, their final values, when compared with the initial point of departure, will be in the same direction as in the short-period adjustment.

b. Effect of Investment Holders

If all the would-be government bond holders, at the existing prices and interest rates, are investment holders, then, as the price of government bonds rises, claims to private producer capital will be substituted, causing their price to rise also. The increase in prices will cause r_G and r_C to fall and thus induce substitution in favor of consumer durables. If we were to assume that the demand for money were inelastic with respect to changes in r_C, this would be the full extent of the adjustment. However, if the demand for money (16a) is elastic with respect to changes in r_C, then the private community, in an attempt to increase its real cash balances, will cause the price level to fall and therefore partially offset the initial movements in prices and interest rates.

In the long run, the fall in r_C will induce producers to increase their capital coefficient and the size of their capital stock, so that output increases and r_K falls; P_C falls relative to p since, in the long-period equilibrium, $P_C = p$. It follows from (9) that r_C must also fall in that r_K must fall with an increase in K. The increase in output will have the effect of intensifying the short-period movements. Therefore, in long-period equilibrium, the rates of interest, r_C and r_G, and the general

1. Cf. T. de Scitovsky, "A Study of Interest and Capital," Economica, N.S. 7, No. 27 (August 1940), p. 300, for a generalization of this reasoning.

2. The stationary model is somewhat strained here since any decline in income requires a decline in the absolute amount of wealth (W), which in turn requires negative saving. Negative saving in the long period can proceed only as fast as the rate of depreciation of physical capital. In a dynamic setting with built-in income growth, the adjustment, more realistically, would take the form of a decline in the savings-ratio. Cf. footnote 1, p. 276 above.

price level, p, will fall while substitution takes place in favor of consumer durables. More new investment takes place than would have occurred had the supply of government bonds been increased proportionately to the rise in income.

One of the assumptions of the model is that expectations remain static. However, this would be highly unlikely in the face of a constant supply of money and government bonds since any increase in income would cause a fall in prices. In the real world this would soon be incorporated into expectations. Under these circumstances, the expected own rate of return on money would no longer be zero; it would rise as a result of expected price declines and would intensify the deflation by increasing the demand for money.

c. The Mixed Case

Where the would-be holders of government securities consist of both liquidity and investment holders, the outcome cannot easily be analyzed in general terms. The larger the proportion of investment holders the less the general price level will fall, but the rate of interest on claims to producer capital will fall more in the short period. In the long period, the capital coefficient of producers will rise more. It is not fruitful to generalize beyond this point since the outcome will depend on the specific values of the various elasticities.

3. The Process of Adjustment and the Real Wage Rate

In a linear homogeneous production function, the real wage rate depends only on the ratio of the stock of producer capital to the labor input. Utilizing this fact, some generalizations can be made concerning the changes in real wages in the above analysis. When the investment holders of government securities predominate, causing the stock of producer capital to increase in the long period, the real wage rate will rise. This is true whether the money supply is constant or variable. However, when the money supply is free to vary, the price level is held constant; therefore the money wage rate must, in this instance, rise. In the case where the money supply is constant and all would-be government security holders are liquidity holders, the real wage rate will fall in the long period as the stock of producer capital diminishes. When the money supply is variable and all would-be government security holders are liquidity holders, the real wage rate remains unchanged.

4. Conclusion

If all the difficulties that a deflationary economic policy would encounter in the American economic setting are considered, the most important being the downward inflexibility of the money wage rate, then certain policy prescriptions may be drawn from the model just presented—assuming the rate of growth of the government debt will not be changed. The process of adjustment will be less painful if the money supply is allowed to grow faster than national income and if the rate of interest is allowed to fall.

The model also showed that when the price level is held constant and the money supply allowed to vary, if any of the would-be holders of government securities are investment holders, then the failure of the government coupon bearing debt to grow proportionately to the national product will result in a greater amount of private investment than would otherwise have occurred. This might increase the danger of a "capital shortage" arising. When the money supply is held constant the situation is, in general, ambiguous and depends on the actual elasticities and the relative numbers of the different would-be holders of government securities.

The exact path of adjustment depends on the distribution of potential government security holders between liquidity holders and investment holders and also, in the real world, on changes in the actual composition of the types of government securities.

I. The Model and the Process of Adjustment: Concluding Remarks

It has been shown how adjustment to changes in the producer capital coefficient and the national debt may be achieved in the short run by changes in the valuation of various claims to wealth (particularly claims to producer capital) and changes in the wealth portfolio of the private community. Long-run equilibrium was attained by changes in the capital coefficient of producers induced by changes in the rate of interest on claims to producer capital. [1]

While the analysis presented is classical in many respects, it differs from the classical model in the emphasis it places on the influence of existing stocks of assets (the result of past saving) on economic behavior. A further difference is the important role assigned to changes in the market valuation of claims to producer capital in achieving short-period equilibrium.

In the specific analysis of a wave of capital-saving innovation serious doubt was cast on the ability of the economy to maintain both full employment and stable consumer prices if wages were inflexible in a downward direction.

The possibility of monetary policy in the conventional sense did not enter into the model since a separate banking sector was not distinguished. However, the model did focus attention on a frequently neglected aspect of monetary policy, namely, the direct asset effects of changes in the supply of government securities on the wealth portfolio of the private community. The analysis implicitly called attention to the fact that the type of government security issue (whether it is a liquidity or investment substitute) has an impact on the process of adjustment apart from its effects on the banking system.

---

1. Perhaps more realistically one should also add changes in the availability of capital funds in that the nominal rate of interest on capital funds may not indicate the relative scarcity of these funds, which the rate of interest is assumed to do in the model.

The general approach enables attention to be focused on the economic implications of the recent fall of the producer capital coefficient and of the decline in the ratio of national debt to national product in the United States.

J. The Wealth-Income Model and Recent American Experience

Can the economic model presented above be applied to explain some of the features of recent American economic experience? Before attempting to answer this question, it is perhaps useful to review some of the more important aspects of economic reality that have been omitted from our model. The model has not given separate consideration to the role of business saving;[1] it has not taken into account the possibility of economic effects of changes in the distribution of income; it has merged the banking sector into the private sector and therefore has eliminated most of the conventional aspects of the impact of monetary policy on the economy; and the role of uncertainty has not received a systematic treatment.

The model indicates that the effects of capital-saving innovations and of a decline in the ratio of the national debt to the national product will largely reenforce one another, provided that the money supply is adjusted so that it will always satisfy the demand for cash balances. Capital-saving innovations will, by definition, tend to lower the capital coefficient of producers and lower the rate of interest by reducing the demand of producers for capital goods per unit of output. However, a decline in the rate of growth of government debt will cause a fall in the rate of interest as a result of increased demand for existing income-yielding assets. This will tend to raise the producer capital coefficient, both through the lower cost and through the increased availability of capital funds. The net effect of these two disturbances on the capital coefficient depends, among other things, on the relative magnitudes of the disturbances and on how factor prices react. The path of adjustment will also be affected by capacity limitations in the capital goods industries.

The two periods in which the disturbances discussed above played a large role were the decades that followed the two world wars. Empirical evidence indicating the path of adjustment is presented in Tables 8–11.

1. 1919–1929

In the period following World War I, total government debt fell from 38 per cent of NNP in 1919 to 31 per cent in 1929 (Table 8). Federal debt actually rose during this period, but the long-term portion of it (and presumably the part of it most appealing to investment holders of the debt) fell from 21 per cent of NNP in 1919 to 13 per cent in 1929.

The capital coefficient in the twenties remained, on the average, stable, although there is some evidence that it rose in the first portion of the

---

1. Cf. N. Kaldor, "Alternative Theories of Distribution," Review of Economic Studies, Vol. XXIII, No. 61 (1955–56), pp. 94–100, for some of the implications of this factor.

TABLE 8

Government Debt and Demand Deposits as a Percentage of NNP
1919–1929 and 1946–1955

Year	Public Debt, Fed. & St.	Net Federal Debt	U.S. Treas. Marketable Bonds	Currency and Demand Deposits	Currency, Demand Deposits, and U.S. Govt. Short-Term Sec.
	1	2	3	4	5
1919	38.4	20.6	21.4	26.4	30.6
1920	34.7	19.3	18.9	30.7	30.7
1921	44.4	27.8	24.2	35.0	35.0
1922	45.7	31.9	24.0	34.5	34.5
1923	37.6	30.4	20.7	30.0	30.0
1924	37.9	38.4	20.2	30.0	30.0
1925	36.5	41.4	20.2	30.5	30.5
1926	33.5	42.3	18.9	29.0	29.0
1927	34.0	44.9	17.4	29.7	29.7
1928	33.6	45.6	14.7	30.5	30.5
1929	31.3	44.9	12.7	29.1	29.1
1930	-	-	-	31.9	31.9
1945	-	-	-	46.4	-
1946	123.1	116.3	60.4	53.2	57.7
1947	109.0	102.4	54.1	49.7	54.7
1948	96.6	89.9	46.3	44.8	49.5
1949	99.1	91.5	43.0	45.2	49.6
1950	90.5	82.7	35.6	41.7	45.4
1951	79.3	71.7	25.3	37.6	46.6
1952	77.3	69.3	24.8	37.7	45.0
1953	76.2	67.7	23.0	36.9	51.4
1954	79.4	69.4	24.6	37.7	50.9
1955	73.8	63.3	22.6	36.3	45.3
1960	65.2	52.2	17.6	31.2	48.7

Sources:

Columns 1–3: Net National Product from 1919–1929, Goldsmith, Vol. III; 1946–1955, U.S.D.C., National Income Supplement 1954 and Survey of Current Business. Government Debt of Columns 1–2, Economic Almanac, 1956, p. 99; Column 3, 1919–1929, Historical Statistics, p. 305; 1946–1955, various issues of the Federal Reserve Bulletin.

Columns 4–5: Currency outside the banks and adjusted demand deposits, Economic Almanac, 1956, pp. 78–79; Marketable U.S. Govt. securities less than one year to maturity less the holdings of the Federal Reserve Banks, Banking and Monetary Statistics, p. 511.

decade and then fell somewhat.[1] This stability in the aggregate conceals the fact that the capital coefficient in manufacturing and mining fell substantially during the twenties, which means that the coefficient in other sectors must have been rising.[2]

While it is not possible to determine whether the money supply was actually equal to the amount demanded as determined in the model, the fact that the ratio of currency and demand deposits, and also short-term Government securities, to national product remained relatively constant (Table 8), and that prices also remained stable throughout the period (Table 12), indicates that this assumption is not unrealistic.

The model would predict that the decline in the government debt would bring about a) a fall in interest rates; b) a rise in the ratio of saving through consumer capital; c) a rise in the ratio of saving through claims to producer capital; and d) in the long run, a rise in the producer capital coefficient.

The prices of securities rose throughout the period, though not evenly, and the yields of securities fell (Table 9). The ratio of saving through consumer capital rose significantly (Tables 10 and 11). The net ratio of saving through securities by the personal sector to disposable income (i.e., new securities sold to the personal sector net of borrowings against securities) was slightly below the period averages. If United States Government securities are excluded, then the ratio of saving through private securities is significantly above the period averages; the same is true if borrowings against securities are excluded (Table 11).[3] It appears that both the substitution effects and the short-run adjustment mechanism utilizing the "wealth effect," caused by changes in the valuation of asset stocks relative to the general price level, were operating.

Thus far the empirical evidence seems to support the conclusions of the model. The one fact that is difficult to reconcile with the model is the stability of the producer capital coefficient throughout the twenties (Table 1). One possible explanation is that, although the lower interest rates and the greater availability of capital funds did, in fact, encourage a higher capital coefficient, this was offset by capital-saving innovations in manufacturing and mining.[4] It is also suggestive that, while the price of capital goods fell both absolutely and in relation to wage rates (Table 12), the latter increased at a rate significantly slower than the average productivity of labor. It is possible that the net marginal revenue product of labor was rising relative to that of capital, and therefore, to a certain extent, offset the effect of the decline in the interest rates and in the price of capital goods. While this explanation is in accord with the existing evidence, it must be regarded as

---

1. Cf. S. Kuznets, op. cit., pp. 82-88 and pp. 122-123.
2. See above, p. 255
3. See R. Thorn, Appendix 32.
4. See footnote 1, p. 255.

TABLE 9

Security Prices and Yields, 1919-1929, 1946-1956

Year	Yields on U.S. Govt. Long-Term Bonds	Yields on Corporate Bonds	Dividend Yields of Common Stocks	U.S. Bond Prices[1]	Corp. High Grade Prices[2]	Index of Common Stock Prices
				(Price per $100 Bond)		(1926 = 100)
	1	2	3	4	5	6
1919	4.73	6.27	5.75	86	82	71
1920	5.32	7.08	6.13	88	75	64
1921	5.09	7.04	6.49	97	77	52
1922	4.30	5.95	5.80	96	86	68
1923	4.36	6.04	5.94	99	85	69
1924	4.06	5.80	5.87	102	87	73
1925	3.86	5.47	5.19	104	88	90
1926	3.68	5.21	5.32	108	90	100
1927	3.34	4.97	4.77	108	92	118
1928	3.33	4.94	3.98	105	92	150
1929	3.60	5.21	3.45	109	89	190
					(1935-39 = 100)	
1946	2.19	2.74	3.75	105	123	140
1947	2.25	2.86	5.06	104	122	123
1948	2.44	3.08	5.87	101	118	124
1949	2.31	2.96	6.82	103	121	121
1950	2.32	2.86	6.51	103	122	146
1951	2.57	3.08	6.29	99	118	176
1952	2.68	3.19	5.55	97	116	188
1953	2.93	3.43	5.51	94	112	189
1954	2.53	3.16	4.70	100	117	227
1955	2.80	3.25	3.93	96	114	300
1956	3.05	3.57	3.89	93	109	342
1960	4.01	4.73	3.46	86	95	414

1. Old series
2. Rounded to nearest dollar

Source:

Column 1: 1919-1929: Historical Statistics, p. 280; 1946-1955: Various issues of the Federal Reserve Bulletin.

Column 2: 1919-1929: Banking and Monetary Statistics, p. 468; 1946-1956: Business Statistics, 1954 and the Survey of Current Business, various issues.

Column 3: Same as Column 2.

Columns 4-6: 1919-1929: Historical Statistics, p. 281; 1946-1954: Business Statistics, 1955, pp. 97, 101.

TABLE 10

Personal Saving and Saving through Consumer Durables as a Percentage of Disposable Income

Year	Personal Saving (U.S.D.C.)[1]	Personal Saving (Goldsmith)[2]	Consumer Durables Purchases (U.S.D.C.)	Consumer Durables Saving (Goldsmith)
	1	2	3	4
1919	-	14.1	-	.66
1920	-	8.9	-	.69
1921	-	2.2	-	.51
1922	-	10.7	-	1.52
1923	-	14.3	-	3.16
1924	-	12.3	-	2.64
1925	-	14.8	-	3.65
1926	-	13.2	-	3.52
1927	-	13.2	-	2.20
1928	-	7.8	-	2.17
1929	16.2	13.9	11.1	2.36
1946	17.9	14.1	9.9	4.0
1947	14.6	12.0	12.2	5.4
1948	17.2	14.2	11.8	4.5
1949	16.6	11.9	12.5	4.7
1950	19.7	-	13.9	-
1951	19.8	-	12.0	-
1952	19.0	-	11.2	-
1953	19.9	-	11.9	-
1954	18.6	-	11.6	-
1955	19.3	-	13.2	-
1960	19.1	-	12.6	-

1. Includes gross purchases of consumer durables.

2. Includes purchases of consumer durables net of depreciation.

Source: National Income Supplement, 1954; Survey of Current Business, various issues, and R. Goldsmith, Vol. I.

TABLE 11

Selected Components of Personal Saving as a Percentage of Total Personal Disposable Income
Social Accounting Concept: Current Values

	1897–1908	1909–1914	1915–1921	1922–1929	1930–1933	1934–1938	1939–1945	1945–1949	Weighted Averages[1]	Unweighted Averages[2]
1. Residential Real Estate	1.8	2.4	1.0	4.2	-1.6	-0.9	-0.0	2.1	1.3	1.4
2. Consumer Durables	1.5	1.4	.6	2.7	-2.1	-1.2	0.2	4.8	1.7	1.3
3. Total Consumer Capital	3.3	3.8	1.6	6.9	-3.7	0.3	0.2	6.9	3.0	2.7
4. U.S. Government Securities	-0.0	0.0	3.6	-1.6	1.0	0.3	7.5	0.2	2.0	1.1
5. State and Local Government Securities	0.3	0.4	0.7	0.6	0.7	-0.4	-0.1	0.3	0.3	0.3
6. Other Bonds	1.7	1.3	2.1	2.2	0.3	-0.9	-0.7	-0.2	0.5	0.9
7. Stock	2.6	2.2	2.3	3.1	1.1	0.5	0.5	1.0	1.5	1.9
8. Mortgages	0.3	1.1	1.6	1.0	-0.2	-0.1	-0.0	0.7	0.5	0.5
9. Total Securities	4.9	5.0	10.3	5.3	2.9	-0.6	7.2	2.0	4.8	4.7
10. Borrowing against Securities	0.4	0.2	0.3	0.8	-2.3	-0.3	0.4	-0.4	0.0	0.1
11. Total Net Saving	10.9	10.4	13.3	12.8	0.4	5.7	22.5	13.4	13.5	11.5

1. Average product of period total products.

2. Average product of annual products.

Source: R. Thorn, Appendix 31.

TABLE 12

Year	Index of Finished Good Prices	Index of Producer Good Prices	Index of Average Hourly Earnings in Manufacturing Including Overtime	Index of Output per Man-Hour in Manufacturing
	1	2	3	4
1919	100	100	100	100
1920	116	114	117	107
1921	103	90	108	122
1922	97	79	102	136
1923	99	83	108	133
1924	99	81	115	140
1925	101	79	115	151
1926	103	76	115	156
1927	100	73	115	158
1928	99	74	117	167
1929	99	77	119	173
1947	100	100	100	100
1948	108	109	109	104
1949	105	114	113	109
1950	107	117	119	118
1951	117	129	128	118
1952	116	131	135	119
1953	115	133	143	123
1954	115	134	146	120
1955	116	138	152	129
1960	157[p]	166	185	-

[p] Preliminary.

Sources:

Column 1: 1919–1929: Consumer price index, 1947–1955: U.S. Statistical Abstract, 1956, p. 313.

Column 2: 1919–1929: Wholesale price index of Metals and Metal Products; 1947–1955: U.S. Statistical Abstract, 1956, p. 313.

Column 3: 1919–1929: Historical Statistics, p. 219; 1947–1955; Economic Almanac, 1956, pp. 240–241.

Column 4: 1919–1929: Historical Statistics, p. 67; 1947–1953; U.S. Statistical Abstract, 1956, p. 220.

speculation on a question to which it may never be possible to give a definite answer. An alternative explanation may be that the production function was of a type which permitted the marginal productivity of capital to fall while the average productivity remained unchanged.[1]

2. 1946–1955

Following World War II public debt fell from 123 per cent of NNP in 1946 to 79 per cent in 1954. United States Government marketable bonds fell from 60 per cent of NNP in 1946 to 23 per cent in 1955. The capital coefficient of producers fell substantially during the war, rose between 1946 and 1949, and fell again between 1949 and 1952. Due to insufficient data on the subject it is not possible to make any definite statement about the postwar trend in the producer capital coefficient; however, this deficiency of material should largely be remedied by the forthcoming capital studies of the National Bureau of Economic Research.

The money supply during most of this period almost certainly equaled, if not exceeded, the amount desired by the private sector, so again our two disturbances tended to reenforce one another, with the exception noted above.

The sharp fall in the capital coefficient which occurred during the war, and which continued into the postwar period, is regarded as evidence that substantial capital-saving innovations had taken place, even though the low values realized during the war were undoubtedly caused partly by full and overfull utilization of capacity, and perhaps by some changes in the composition of output.

Both the decline in the capital coefficient and in the ratio of national debt to NNP have led to a decline in interest rates. However, the course of events was somewhat more complicated. Bond prices, on the whole, fell during the postwar period, while stock prices remained relatively stable until 1949 and then rose steadily until 1956. Correspondingly, the yields on bonds during the period rose while the yields on stocks fell (Table 9). As is well known, the price of Government bonds was supported during the war and postwar period by the Federal Reserve, although after 1951 this support was made more flexible under the terms of the Federal Reserve–Treasury "accord." Without undertaking a full review of the effects of Government bond support by the Federal Reserve, it can be said that the net result of this support was to keep the interest rate on government and private bonds well below the rate that would have prevailed in the type of free market envisioned in our model. It can be further maintained that the price movements in the equity market gave a more accurate indication of the relative demand for and supply of claims to capital than did those of the bond market.[2]

One of the features of the postwar securities market was the domination of the private bond market by the Government bond market. In the period

1. Cf. F. M. Bator, op. cit.

2. Cf. Scitovsky, op. cit., p. 300, for a theoretical discussion of this phenomenon.

following World War II, the spread between the interest rates of Government bonds and of private bonds decreased, and the spread between private bonds and common stocks increased (Table 9). In the twenties there were six divergent movements between annual United States Government bond prices and corporate high grade bond prices and four divergencies between the latter and common stock prices. In the period following World War II there was only one divergency between Government and private bond prices and six divergencies between private bond prices and common stocks (Table 9). It may be maintained that the long-run trend in the market rate of interest which the model would lead one to expect was actually occurring in the only market that was comparatively free to respond to the relative demand for and supply of capital funds.[1] In fact, it may be argued that the low rate of interest in the private capital market, influenced strongly by the Government securities market, was one of the factors responsible for the "capital shortage" that arose.

The rate of personal saving through consumer durables, while at first high due to the backlog demand of the war, seemed to be continuing at a rate certainly equal to, if not exceeding, that of the twenties.

The fact that the "real" wealth coefficient of the non-farm sector was significantly below the levels historically "desired" by this sector (assuming that the realized historical values fluctuated about the "desired" values), would, from our model above, suggest that the savings-ratio should be above its historical value as households again attempted to restore their "desired" wealth coefficient. The high rate of personal saving that prevailed in the postwar period supports this conclusion.

The period also gave evidence of what was analyzed as a "capital shortage," brought about by the desire of producers to utilize the concentrated wave of innovations that were introduced during the war and by the high demands placed on the capital goods industries by the national defense program. As a result of the downward inflexibility of wages and an unwillingness to tolerate a certain amount of deflation and unemployment, the adjustment to this "capital shortage" was largely in the nature of a price adjustment, as described above. Prices of producer goods rose in relation to those of finished goods (Table 12). Wage rates, however, rose faster than producer goods prices, and also faster than the average productivity of labor (Table 12). Although the data presented are for manufacturing alone, it may be assumed that the same general movements prevailed in the rest of the economy.

The possibility exists that the rise in wages had outrun the rise in the marginal productivity of labor, which (unless producers were willing to

---

1. The equity market, of course, is not independent of the other sections of the capital market, but there is reason to believe that it was perhaps the least affected by the Federal Reserve's support of Government security prices and the control of its monetary policy. See footnote 2, p. 294.

accept lower profit margins) would tend to increase the substitution of capital for labor and provide a strong incentive to introduce labor-saving innovations. In addition, the fact that producers were faced with a rising trend in the prices of capital goods and labor was a strong inducement to invest immediately and to substitute capital for labor. This reaction would tend to intensify the severity of the "capital shortage."[1]

In the face of these facts it seems that the rate of interest, as it affected the cost of capital, had taken a back seat since it had to offset not only the substitution effects caused by the rise in wages, but also the expectation of a rise in the price of capital goods.

If the government had tried to follow a high interest rate policy, private wealth holders might have substituted saving through capital claims for saving through durables. While this latter substitution effect might have eased the process of adjustment by freeing resources for the expansion of the capital goods industries, it could do little in the short run to overcome the problem of long gestation periods in some industries, and little to alleviate specific shortages of capital and labor. The policy that would have seemed desirable in this situation, if the problem was actually one of a "capital shortage," would have been the introduction of some type of selective credit controls which would have permitted expansion of the capital goods industries alone. As a matter of practical policy, this might mean expanding credit in some sectors of the economy while applying strong general monetary restraints elsewhere to mop up the undesirable secondary effects of such a program. In the long run, however, there exists the problem of adjustment to a decline in the capital intensity of production and to the diminished relative size of government debt.

K. A Look Ahead

The model presented is relatively simple but is capable of being widely generalized and extensively refined. The same remarks apply to the statistical technique employed. The value of these generalizations and refinements depends on whether it is considered worthwhile to look upon the economy from the viewpoint of changes in stocks of assets, as was done here, and on whether changes in the capital intensity of production constitute an important systematic variable in problems of long-run adjustment.

The next steps in the generalization of the model would be, first, to integrate the effects of wealth holding and capital intensity of production with the rate of growth of income and, secondly, to account for changes in the distribution of income, particularly between wage earners and profit receivers. I believe identity (18) above offers a clue to the solution of the first problem, and that further sub-sectoring of the private sector by relative income position, e.g., by quartiles, would tend to solve the second

1. There are also some indications that the producers of capital goods may have been more conservative in their expansion programs than other producers in the economy.

problem, since the upper income group includes most of the profit receivers.

Wesley C. Mitchell often pointed out that there is a definite relationship between the economic problems of the day and the type of economic theory that is developed.[1] It was no coincidence that much of the economic theory and the conclusions that were so skillfully set forth by Keynes were arrived at independently by other investigators. It is no coincidence today that so many investigators are emphasizing the role of asset stocks and their composition and distribution in the explanation of economic behavior and analysis of economic growth. The problems at hand have underscored the importance of this approach. There is no doubt that lack of adequate data for statistical analysis on the holding of wealth and the utilization of producer capital is preventing a "break through" in the explanation of this important aspect of economic behavior.

Lastly, one may ask: what light does the present analysis shed on the problems which may confront the American economy in the future? In the empirical section of this paper it was shown that a sharp fall in the producer capital coefficient has occurred during the present postwar period. If the ratio of government debt to national product continues to decline, and there is sufficient expansion in the money supply to provide an adequate amount of liquid assets to the private sector, and if the producer capital coefficient should prove interest-inelastic in the short run, the supply of capital funds per unit of output will exceed the demand for these funds per unit of output, and the absolute level of investment will rise sharply, creating an inflationary, or potentially inflationary, situation.

Equilibrium will be restored in the long run as a result of diminishing returns on additions to producer capital, the average producer capital coefficient eventually being raised. However, innovations may occur at a rate sufficient to offset the fall of the marginal productivity of producer capital, in which case the inflationary situation will continue. The more interest-elastic the producer capital coefficient, the smoother the process of adjustment will be, and less severe changes in the rate of saving and investment will be required to attain a new equilibrium.

If the effects of downward inflexibility of wage rates and the possible occurrence of "capital shortages" are considered, it may be that the monetary authorities employing the orthodox concepts of monetary policy will be faced with the Scylla of inflation and the Charybdis of unemployment without any safe middle course. In this event the possible asset effects of government debt of the "proper variety" may permit a policy to be devised which would steer a more satisfactory middle course by allowing high interest rates to choke off investment demand, while at the same time absorbing potential capital funds in government securities. These asset effects would be independent of those operating through the banking system

1. W. C. Mitchell, Lecture Notes on Types of Economic Theory, 2 vols. (New York: Augustus M. Kelley, 1949).

and would affect many non-bank financial intermediaries which are partially insulated from conventional monetary policy.

I shall end this essay with a question. Should the desire of the community to hold a greater amount of wealth assets per unit of income than is currently available at a given rate of interest be allowed to force the economy into undertaking investments of very low productivity per unit of producer capital? The desire of the community that leads it into these low productivity investments may be primarily the desire to acquire claims to wealth, and only secondarily to the income which they yield. The motivation of the community is somewhat analogous to the boy who encourages his mother to buy a particular kind of breakfast cereal so he can save the box top even though he is not particularly anxious to eat the cereal.

The government has the alternative of providing the private community with income yielding claims to wealth, the supply of which has the effect of maintaining a higher average productivity of private producer capital than would otherwise prevail. There are several reasons why this policy may be desirable.

It is possible that the marginal productivity of certain types of government investment is higher than that of private investment, e.g., roads and large-scale power projects. The public welfare may be improved more by government investment in schools, hospitals, parks, etc., than by an equivalent amount of investment in private capital.

In a country with limited natural resources, it may be desirable, from the point of view of long-run maximization, to postpone investment with low capital productivity which exhausts the stock of natural resources until innovations appear which will raise the productivity of capital and maximize the satisfaction that a given stock of resources may yield.[1] It is implicit in this reasoning that the government deficit necessary to implement a high private capital productivity policy will be used to finance government expenditure of a type which will utilize a smaller amount of natural resources than would otherwise be the case.

Lastly, there exists the consideration that, should the rate of innovation be sporadic, it would be a wise policy to maintain a high average productivity of capital so that in periods when the rate of innovations slows down the level of investment in producer capital may be sustained. This may be achieved by a falling rate of interest until either the rate of innovation increases, or, if this does not seem likely, until policies to combat the impending "stagnation" can be developed and implemented.

There is an upward limit on how far the government may restrict private investment to those investments with a higher producer capital

---

1. Such a conclusion, of course, raises all the problems of the intertemporal and interpersonal comparisons of welfare. There is also the problem that the stock of what are considered natural resources of a country are not independent of changes in technology.

productivity, since the method by which the government does this (e.g., raising the rate of interest on private capital funds) also lowers the rate of growth of income (which need not necessarily be the case if the output of government capital formation were included). The upward limit is set by the rate at which diminishing returns accrue to additions to the capital stock and by the politically necessary rate of income growth.

It is not at all obvious a priori that the objectives of income growth, welfare, or cyclical stability will be maximized over a period of time by a national debt which has a zero or negative rate of growth.